A genomic investigation of *Meloidogyne javanica*: genome assembly, comparative analysis, and virulence

being a Thesis submitted for the Degree of PhD

in the University of Hull

by

Michael Robert Winter

MSc, BSc (Hons)

September 2024

0.1 Declaration of Authorship

I declare that this work was undertaken by myself and is my own. The thesis received revision from my supervisor David H Lunt. Chapter 2 received contributions from collaborators:

Chapter 2

Published in **Winter, M.R.**, Taranto, A.P., Yimer, H.Z., Coomer Blundell, A., Siddique, S., Williamson, V.M. & Lunt, D.H. (2024) Phased chromosome-scale genome assembly of an asexual, allopolyploid root-knot nematode reveals complex subgenomic structure. PloS One, 19(6), p. E0302506.

MRW was responsible for conceptualization, data curation, formal analysis, investigation, methodology, project administration, software, validation, visualisation, writing – original draft, and writing – review & editing. **APT** was responsible for conceptualization, data curation, formal analysis, investigation, methodology, project administration, software, supervision, validation, and visualisation. **HZY** was responsible for investigation and methodology. **ACB** was responsible for investigation. **SS** was responsible for conceptualization, funding acquisition, methodology, resources, supervision, and writing – review & editing. **VMW** was responsible for conceptualization, investigation, methodology, resources, visualisation, writing – original draft, and writing – review & editing. **DHL** was responsible for conceptualization, funding acquisition, methodology, project administration, resources, validation, writing – original draft, and writing – review & editing.

0.2 Acknowledgements

I would like to acknowledge the ongoing and unwavering support provided by the people in my life throughout the generation of this work.

First and foremost, my wife, Pelina. Your unwavering patience and constancy throughout this process has been invaluable, without which I am certain I would have been unable to attain what I have achieved.

To my son, Lucas. I expect that your memories of Daddy "writing his book" will only become more distant as you grow, though with your unbridled curiosity you will no doubt pull this thesis from the shelf at one or more points in your life. I hope that this work inspires you, son, as you inspire me.

To my family, who have supported me over the many years since I decided to run in this direction. You told me I was capable of greater things, and gave me a place to rest until the time came when I was able to do them. I am forever grateful.

To my friends, both inside of academia and out, thank you. From listening to my stresses and failures, to celebrating my successes, or even just being an ear to vent frustrations and fanciful ideas to, you all helped immensely.

I would like to thank my supervisors, collaborators, and colleagues, including all members of EvoHull, specifically Dave Lunt, Africa Gómez, Adam Taranto, and Graham Sellers. Your formal and informal mentoring - whether purposeful or incidental - has given me a broader suite of skills than I could otherwise have developed, and your wisdom has guided me through this process.

Finally, I would like to acknowledge those who were here when I began this journey, but were unable to see it completed. Thank you for everything. I hope you are proud.

0.3 Contents

0.4 List of Figures

0.5 List of Tables

0.6 Reproducibility statement

Here we provide a reflection on reproducible practices in our field, and how we have considered and attempted to adhere to and build upon them.

The extent of open research practice in the field

The study of genomics and bioinformatics is highly suited to the application of open research practices, both in the publication and sharing of data and the availability and accessibility of methods and bioinformatic workflows (Baykal et al., 2024). Alongside this many bioinformatics and genomics manuscripts are published in open access journals. At the start of this research awareness and adherence to FAIR practices was becoming more widespread, however adoption of these mantras was not commonplace (Wilkinson et al., 2016).

Free and public access to genomic data provides opportunities for further validation of results and inclusion of previously published data in novel analyses. Uploading of data to these databases is essentially obligatory in the field. The caveat to this is in areas of genomics studying and utilising human data, where availability and access can be tightly controlled due to data protection legislation (O'Doherty et al., 2021). Since 1984 it has become routine for sequence data to be uploaded to the International Nucleotide Sequence Database Collaboration (INSDC), at either NCBI Genbank or Sequence Read Archive (Arita et al., 2021). This has not changed during the development of this thesis and research.

Many studies in this field make their bioinformatic methods open access, although in varying states of practical reproducibility. Availability of established bioinformatic workflows in a format that is deployable by future researchers drastically reduces the time and cost required to perform analysis independently, while also enabling comparison to other studies through the application of similar or parallel methods using the same workflow (Wratten et al., 2021; Ziemann et al., 2023). Many bioinformatic projects utilise Github and Zenodo to ensure long term accessibility of methods (Ram, 2013; Turkyilmaz-van der Velden et al., 2020; Braga et al., 2023). Methods used to ensure the openness and reproducibility of bioinformatic methods have come far since the beginning of the work presented in this thesis, aided by advancements in computational methodology such as environment management, version control, public repositories, and workflow managers or containers. The use of workflow managers to perform and standardise complex bioinformatic pipelines was far from widespread, with those existing and publicly available still being opaque to their

exact methodology, and offering little modularisation or adaptation potential other than that provided by the original package.

Considerations of openness and transparency

In accordance with the standard and expectations of the field, all raw sequencing data generated throughout this thesis will be made publicly available - where it is not already - on NCBI Genbank and the Sequence Read Archive for free use by future researchers (Arita et al., 2021).

All bioinformatic methods in this thesis fall under tiers of automation and availability. Where possible, all complex bioinformatic methods are contained in automated workflows managed by *snakemake*, a Python based workflow manager that allows modularisation, version control, reporting, and robust reproducibility (Köster & Rahmann, 2012b). In cases where an analysis required more oversight or was not suitable for automation, we use *Jupyter* notebooks. *Jupyter* notebooks allow reproducible bioinformatics by providing a step-by-step workflow, that although unautomated, is highly transparent and easy to use, requiring any researcher attempting to replicate the method to follow preset, ordered instructions using predefined code (Kluyver et al., 2016). In rare cases where a task was too small to require a notebook, or required application of a programming language or method incompatible with *Jupyter* notebooks, we provide the code in a script. All programmatic and computational methods are recorded thoroughly, and are reproducible, as well as malleable to be applied to other analyses following alteration of inputs and parameters. Wherever possible, all computational methods have been developed and designed with best practices for programming in mind (Stodden et al., 2016; Ziemann et al., 2023). Chapter 2 of this thesis was published as a manuscript in an open access journal (Winter et al., 2024), and it is our intent to do this with any further publications. We maintain that the framework of reproducibility displayed here is at the cutting edge of what is considered reproducible in both genomics and bioinformatics, and we hope that through leading by example future researchers will utilise and emulate the reproducible methods and adherence to open research practices presented here.

0.7 General Abstract

The *Meloidogyne incognita* group (MIG) of nematodes, including *Meloidogyne javanica*, contains some of the most destructive crop pathogens worldwide. Genetic resistance, such as through the *Mi-1* gene, is used to limit *M. javanica* infection, but resistant strains like VW5 have emerged, increasing the threat posed by this species. Understanding of the molecular mechanisms driving speciation and adaptation in the MIG is limited by outdated genomic resources. We generated a chromosome-scale, fully annotated genome assembly of *M. javanica*, revealing an allotetraploid genome composed of two diploid subgenomes (A and B). Subgenome B shows evidence of chromosomal fission, despite considerable synteny with subgenome A. This assembly provides a valuable resource for studying allopolyploid genome evolution and pathogenicity mechanisms in the *Meloidogyne* genus. Using the novel *M. javanica* assembly as a reference, we phased and annotated genomes of other MIG species, as well as *Meloidogyne luci*. We performed synteny and phylogenetic analyses across these species at a subgenome scale, finding conserved synteny and collinearity between subgenomes and species. No clear evidence of large-scale dominance was found, and instead we reveal a more nuanced picture whereby dominance seems to be balanced, occurring strongly at individual loci but with no clear subgenome scale trend. Phylogenetic analysis indicates that subgenomes are more closely related to orthologous subgenomes in other species than to homoeologous subgenomes within the same species, suggesting a shared ancestry. We also find that *M. luci* is allopolyploid, with subgenomes orthologous to those of the MIG. This suggests that many clade I *Meloidogyne* species are likely allopolyploids, with shared parent species. This highlights the utility of this group as models for studying the evolutionary dynamics of allopolyploid genomes. Finally, we investigate a genomic region involved in resistance evasion of the *Mi-1* gene in the VW5 lineage of *M. javanica*. We discovered a ~650 Kbp deletion on one subgenomic copy, encompassing the single complete copy of the *Cg-1* region. The loss of key transcripts from this deletion suggests a possible mechanism for resistance evasion and adaptation through gene loss, where the absence of a secreted effector may enable the nematode to evade host defences. Overall, this work provides new genomic resources and insights into the evolutionary dynamics of subgenomes, selection pressures, and resistance adaptations, helping to advance research and control methods for these globally significant plant pathogens.

Chapter 1: General Introduction

Michael R Winter[1]

[1] School of Natural Sciences, University of Hull, Kingston-upon-Hull, UK.

Correspondence: mrmrwinter@gmail.com

1.1 Nematoda

Nematoda is a phylum of animals known colloquially as "roundworms". Nematodes are thought to have been first recorded by Hippocrates, who described ascarids - small intestinal unsegmented worms occasionally found in the faeces of humans and animals (Moulé, 1911). The term "nematode" - meaning "thread, slender with cylindrical shape" - was later coined to describe both ascarids and strongylids - hookworms, found infecting other parts of an animal - and was first used in 1803 as "nematoides", coalescing in 1839 as "nematodes" (Hugot et al., 2001; Rudolphi, 2008).

The majority of animals on Earth are predicted to be nematodes by individual counts, and the phylum is suggested to account for ~1% of the biomass of life on earth (Bar-On et al., 2018)) with an estimated ratio of nematodes to humans of 57 billion to one, respectively (van den Hoogen et al., 2019).

1.1.1 Description and biology

Nematodes are unsegmented, pseudocoelomate ecdysozoans estimated to have diverged into a distinct phylum during the Ediacaran period around 650 million years ago (Lee, 2002; Shu et al., 2014).

The nematode life cycle generally adheres to the following process, consisting of six stages. First is the egg, followed by four larval instar stages (J1, J2, J3, and J4), each demarcated by a moult, before finally the J4 moults into an adult (Wharton, 1986a).

They exist in a vast range of sizes, from *Placentonema gigantissima*, a nematode species that reaches 8.4 metres long and parasitises the placenta of sperm whales (Gubanov & Others, 1951), to microscopic nematode species such as *Stegelletina lingulata*, which reach ~500 microns and are found in decomposing vegetable matter (Abolafia & Shokoohi, 2017).

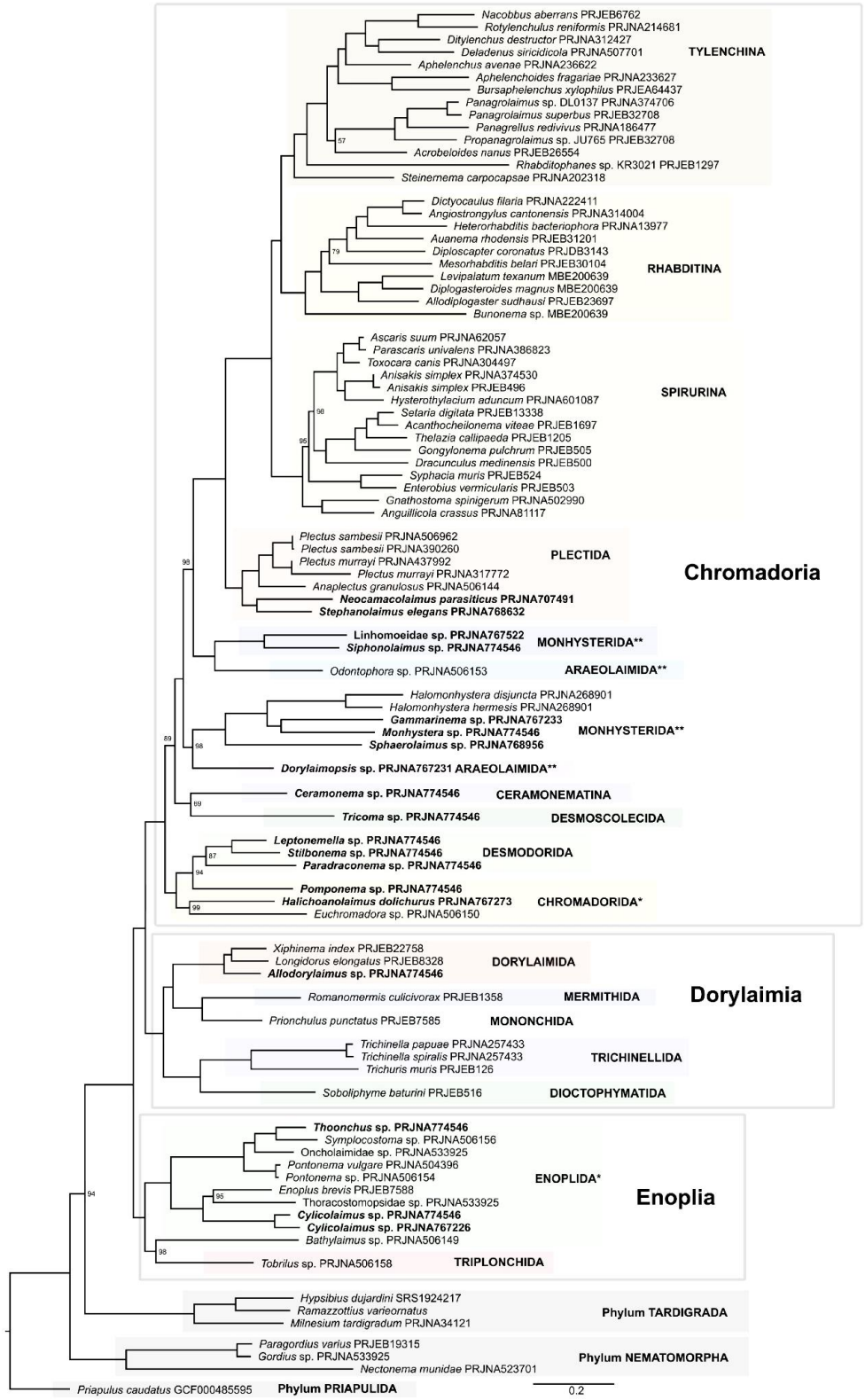


Figure 1.1: Phylogeny of Nematoda (Ahmed et al. 2022; Figure 6). Inferred from 416 orthogroups. Only bootstrap support values <100% are shown. Paraphyletic orders are marked with *, polyphyletic orders are marked with **.

1.1.2 Nematode classification

The history of phylogeny and classification of Phylum Nematoda is convoluted, with the most recent classification system based on molecular evidence defining 3 classes, 8 subclasses, 12 superorders, 32 orders, 276 families, and 3,030 genera, encompassing a total of 28,537 named species, although this number has likely expanded since this classification system was established (Hodda, 2022). Estimates of total species richness for this phylum exceed 1 million (Lambshead & Boucher, 2003; Blaxter et al., 2005). Robust phylogenies of Nematoda have been produced, revealing the major classes and orders (Figure 1.1; (Blaxter & Koutsovoulos, 2015; Ahmed et al., 2022).

1.1.3 Behaviour and life history

Nematode life history ranges from entirely free-living to obligate parasitism of both animals and plants, although as ecdysozoans, all exhibit several life stages demarcated through periodical shedding of the outer cuticle. In nematodes, these stages typically include an embryonic stage, four to five larval stages - often referred to as juvenile, followed by the number of sheds, ie, J1, J2, etc - and an adult stage, though given the diversity seen in a group as large and diverse as Nematoda this is a simplification (Lee, 2002; Sommer & Streit, 2011).

Sexual systems also vary, from gonochorism to obligate parthenogenesis (Castagnone-Sereno & Danchin, 2014; Van Goor et al., 2023). Sex determination in nematodes can be influenced by both genes and environment. Some species develop strictly male or female individuals, some exhibit a number of hermaphrodites alongside males and females, and some are exclusively hermaphroditic and therefore parthenogenetic (Hodgkin, 1983; Blackmore & Charnov, 1989; Stothard & Pilgrim, 2003). Methods of parthenogenesis also vary between meiotic (automixis) and ameiotic (apomixis). These combinations lead to a complex array of reproductive methods, genomic complements, and genetic diversity across the phylum (Wharton, 1986b). Most species are oviparous, with females (or hermaphrodites) laying many eggs at a time, although some are viviparous, birthing live larvae (Hugot et al., 2001), and some are ovoviviparous, with eggs hatching internally, wherein they consume and eventually erupt from the mother (Chen & Caswell-Chen, 2004). Predators of nematodes include, alongside other nematodes, insects, tardigrades, mites, centipedes, symphylans, and nematophagous fungi (Kiontke & Fitch, 2013).

As is expected for a group as large and diverse as Nematoda, animals within it fill many niches, spanning most conceivable habitats, including terrestrial, marine, and freshwater, as well as within other organisms as parasites. There are extremophilic nematodes that exist in and around undersea hydrothermal vents (Gerlach & Riemann, 1973; Thiermann et al., 2000). Nematodes exist in arid deserts (Freckman et al., 1974, 1975) and tropical rainforests (Porazinska et al., 2012), and have been found in polar regions at both ends of the globe (Holovachov, 2014; Elshishka et al., 2023) and in deep mines, kilometres underground (Borgonie et al., 2011).

1.1.4 Parasitism as a strategy

Given the broadly encompassing nature of nematode life history, behaviour, and environment, the diet of nematodes varies greatly, however parasitism, both facultative and obligate, is a common strategy (Hugot et al., 2001; Davis et al., 2004; Castagnone-Sereno & Danchin, 2014). Some species are free-living as larvae, maturing to adulthood upon infection of a host, whereas some species spend their entire life cycle within the host (Poulin, 2011; Blaxter & Koutsovoulos, 2015; Poulin & Randhawa, 2015).

Nematodes infect both animals and plants, causing reduced vitality, disease, and often death. In humans, it is thought that over 25% of the population is affected by nematodes (Bethony et al., 2006). This number increases to between 50-80% of livestock, with wild populations likely exhibiting similar prevalence (Chavhan et al., 2008; Nouri et al., 2022; Tachack et al., 2022). The damage they inflict upon the animal host can vary. Some species can reside within a host for much of the host's life, inflicting little noticeable damage while others are deadly, an example of which is the hookworm, which is estimated to be responsible for the deaths of over 65,000 humans per annum (Diemert et al., 2008). Another deadly example is *Dirofilaria immitis*, which infects cats and dogs globally, causing heartworm disorder and resulting in death (Noack et al., 2021). Incidental parasitism - where a parasite infects an organism other than the host taxa - can also be deadly. This is often the result of extraintestinal migration of the nematode in the incidental host, which in some cases can enable the parasite to migrate to the brain or central nervous system, causing debilitating disease and death (Walker & Zunt, 2005).

1.1.5 Plant parasitic nematodes

One of the many niches that nematodes fill is as parasites of plants (Davis et al., 2004; Blaxter & Koutsovoulos, 2015). Plant parasitic nematodes (PPN) are typically less deadly to the host, however they cause wide ranging secondary impact through crop loss and agricultural impact. Estimates for the economic impact of PPNs vary and are hard to ascertain with definite certainty, but range from between 80-173 billion dollars (USD) annually worldwide (Jones et al., 2013; Singh et al., 2015; Bernard et al., 2017; Talavera-Rubia et al., 2022). This number does not wholly account for the impact of PPN in the developing world, where in the absence of awareness of PPN and availability of diagnostic and control measures, total crop loss and famine are a frequent outcome of infection (Mitiku, 2018; Mandal et al., 2021).

Many Families of nematodes specialise in plant parasitism, however the most impactful and damaging belong to the family Heteroderidae (Jones et al., 2013). Heteroderidae includes both the cyst nematodes (genera *Heterodera* and *Globodera*), and the root-knot nematodes (RKN; genus *Meloidogyne*) (Williamson & Hussey, 1996; Jones et al., 2013). Parasitism by PPN directly induces morphological changes in the host, resulting in deformed root structure, from where these genera get their name. Species in these groups induce the formation of "giant cells" through injection or secretion of proteins that cause cellular hypertrophy and multinucleation. The nematode then feeds on the increased volume and density of the cytoplasmic content of these giant cells. Infection and feeding is aided by a stylet: a hollow mouth part similar to a hypodermic needle, which is used mechanically to break through the relatively tough wall of the root, as well as feed off of giant cells (Hussey, 1989; Wyss & Grundler, 1992). PPN are motile as J2s, which penetrate the host root and navigate intracellularly to the preferred feeding site - the zone of differentiation - where they moult into adults and become sedentary for the remainder of their life cycle.

1.2 Root-knot nematodes (*Meloidogyne*)

Meloidogyne is a genus of nematodes known as root-knot nematodes (RKN). First identified by Cornu (1879), as *Anguilla marioni*, the *Meloidogyne* genus now contains around 100 species (Chitwood et al., 1952; Jones et al., 2013; Subbotin, Palomares Rius, et al., 2021; Abrantes et al., 2023).

RKN can infect almost all species of flowering plants, including virtually all agricultural crop species, and are highly impactful crop pests (Jones et al., 2013), costing the agricultural industry billions of dollars annually (Bernard et al., 2017). The majority of the most impactful RKN species are endemic to tropical latitudes but *Meloidogyne* species have been found on all continents other than Antarctica (Perry et al., 2009). The most impactful species of this group are *M. hapla, M. arenaria, M. incognita,* and *M. javanica*, the latter three of which belong to the infamous *Meloidogyne incognita* group (Trudgill & Blok, 2001).

1.2.1 Biology

Like all nematodes, RKN are pseudocoelomates, containing no internal skeleton and maintaining turgidity with a cuticle, which is moulted at several stages of its life. Free-living J2s are no bigger than 500 µm and 15 µm wide. Adults are around 1500 µm in length, with females having a highly variable width due to various levels of swelling according to egg production (Mitkowski & Abawi, 2003). Based on examples of *Caenorhabditis elegans* with similar lengths, this will amount to around one thousand individual somatic cells (Gilbert, 2000). RKN possess a stylet, similar to other PPN. A stylet is a sharp tubular appendage that can be protruded from the anterior end of the nematode, and is used to pierce the cell wall of plant cells in order to inject secretions and ingest cell contents (Figure 1.3a).

RKN can thrive in almost all examples of terrestrial environments, and reside in the uppermost one or two feet of soil, whether free-living or sedentary in a host (Olsen, 2000). The life cycle of a RKN is similar to that of most nematodes in that it begins as an egg, transitions through four juvenile larval stages (J1 through J4) with each moult, before moulting into an adult after the J4 stage (Figure 1.4). The variable factor in the life cycle of a RKN is its behaviour, transforming from a relatively free-living juvenile to an entirely sedentary adult (Perry et al., 2009). The egg is released and the J1 proceeds to develop within it. The first moult occurs prior to hatching, beginning the J2 stage whilst still inside the egg. J2s hatch from the egg and are free-living within the tissue of the host, gradually migrating to the vascular cylinder of a root apex in the host (Figure 1.3b). Once there the J2 establishes a feeding site, becomes sedentary, and proceeds to moult three more times to become a mature adult. The worm remains here for the remainder of its life, inducing a gall in the plant tissue (Figure 1.3d) and swelling with eggs, which eventually restarts the cycle (Figure 1.4). The exception of this is males, which remain free-living for their lifetimes and migrate out of the host (Abad et al., 2003).

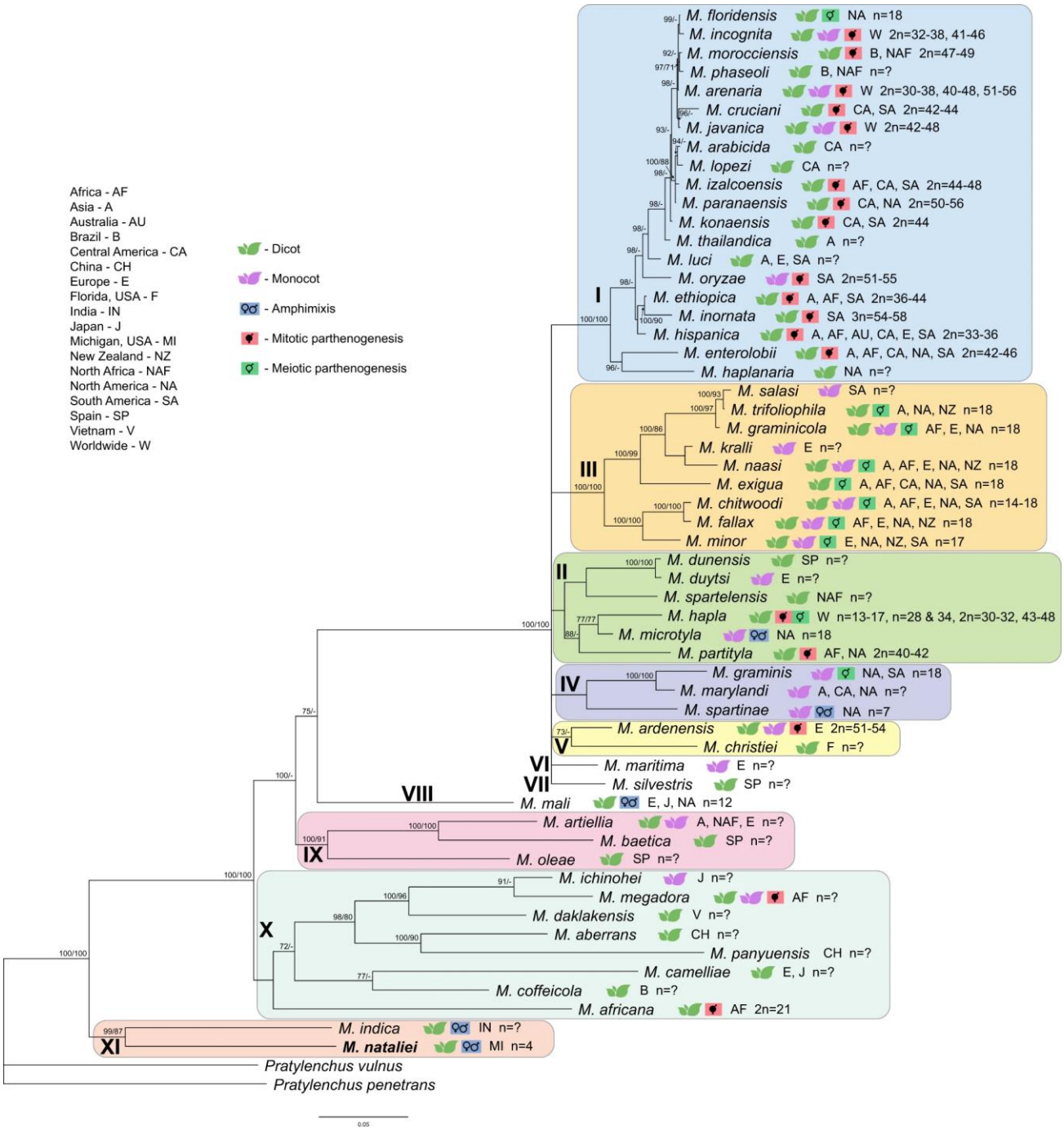


Figure 1.2: Phylogeny of *Meloidogyne* (Alvarez-Ortega, 2017; Figure 4). Inferred from 18S rRNA, ITS1 rRNA, D2-D3 expansion segments of 28S rRNA, COI gene and COII-16S rRNA. Clades are coloured and assigned roman numerals. Clade I can be seen in blue, containing the MIG species among the most distal branches. It can be seen that the majority of clade I species reproduce through asexual methods, usually mitotic parthenogenesis (apomixis).

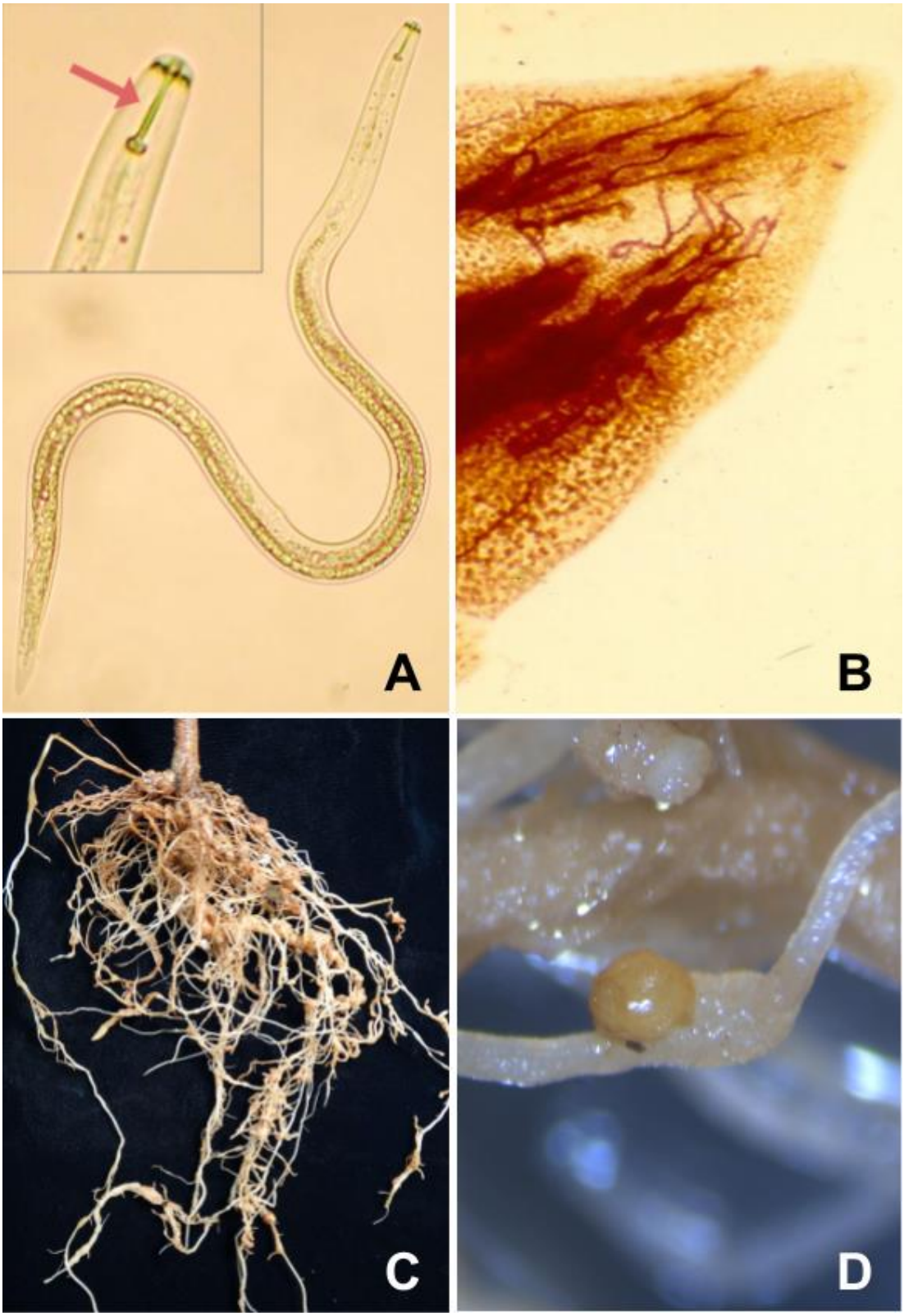


Figure 1.3: A, Plant parasitic nematode. Inset, close up of stylet, labelled with red arrow. **B,** *Meloidogyne arenaria* (J2) in peanut root tip (*Arachis hypogea*). **C**, *Meloidogyne* incognita infection in tomato (*Solanum lycopersicum*). Galling can be seen in roots. **D,** Egg mass of *Meloidogyne incognita* on galled tomato root. Egg masses contain hundreds of eggs, in a gelatinous matrix (Nelson 2022).

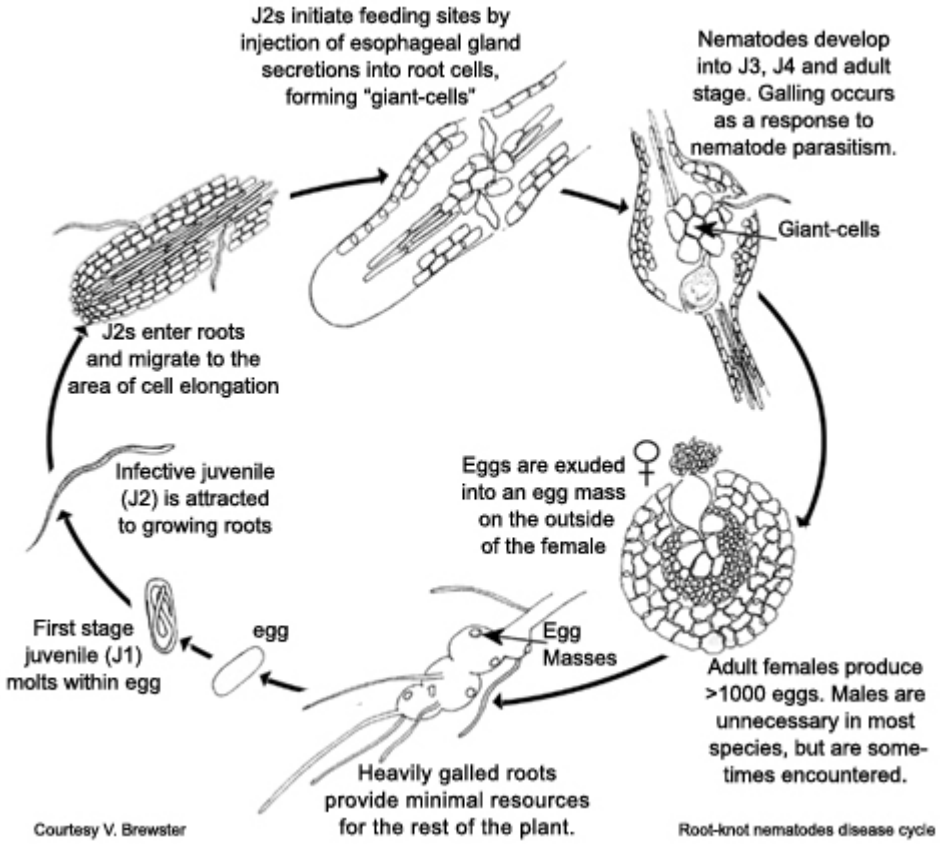


Figure 1.4: Life cycle of a root knot nematode (Al Abadiyah Ralmi et al., 2016)**.** RKN begin life as a singular egg belonging to a larger egg mass containing over a thousand others that recently erupted from a gravid female. The J1 develops within the egg, undergoing its first moult before hatching and becoming a J2. J2s hatch from the egg and are free-living, migrating through the host toward the root tip, before eventually establishing a feeding site and becoming sedentary. In response to effector proteins injected and secreted by the RKN and the plants host defence mechanisms, a gall forms around the feeding site, leading to dysmorphia of the root, ie, a root knot. Whilst here J2s undergo three more moults, becoming an adult RKN, and producing its own egg mass.

1.2.2 Parasitism by RKN

Nematodes of Genus *Meloidogyne* can likely infect all species of angiosperm, including all human agricultural crops (Perry et al., 2009; Jones et al., 2013). Upon establishing a feeding site, the RKN begins inducing the growth of giant cells (GC) in the host (Figure 1.5).

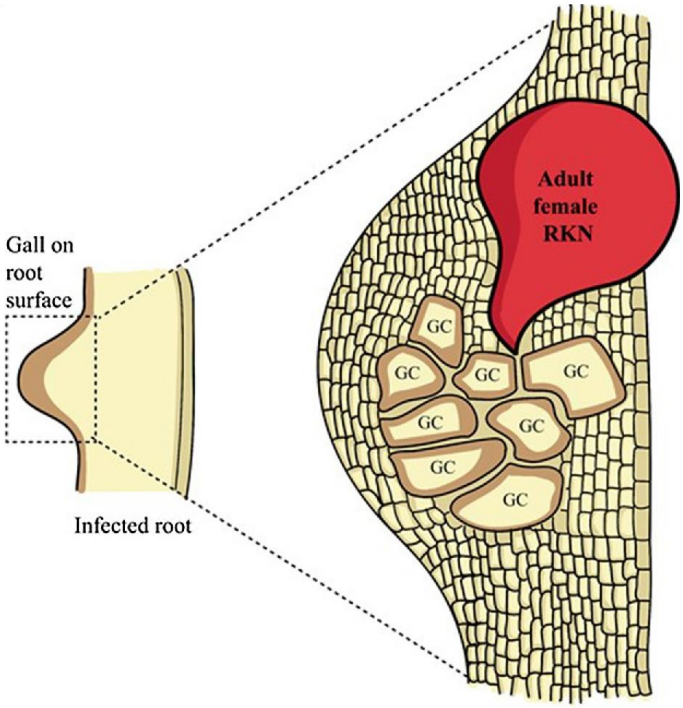


Figure 1.5: Diagram of RKN gall showing position and relationship of nematode and giant cells (Jagdale et al., 2021)**.** Adult female nematodes reside within the gall alongside GCs on which they feed.

GCs are plant cells in the vascular column of the root tip that become engorged with cytoplasm and hypertrophied, multinucleated and up to 300 times the size of a normal cell (Mejias et al., 2019). Within the GC, certain proteins beneficial to the RKN are upregulated, at the cost of fitness to the host. The enlargement of multiple GCs throughout the life cycle of the nematode causes abnormalities in the morphology of the host root, leading to the formation of galls (Figure 1.2D) and "root-knots" (Figure 1.2C). The effect on the plant can be substantial, depending on the extent of infection, ranging from a limited ability to thrive up to reduced yield and host death (Feyisa, 2021).

Host invasion and GC formation are regulated by the release of "effectors" by the RKN; proteins which modify processes in the host cell, such as suppressing infection response and defence mechanisms, enabling the ease in which the RKN can permeate the cell wall or move through the root, or inducing multinucleation and subsequent GC formation (Nguyen et al., 2018; Rutter et al., 2022). These proteins are typically produced in esophageal gland cells and injected through the stylet, although some are secreted directly from the cuticle into the extracellular space of the host. Over 500 secretary proteins have been identified, each contributing to the rich and little understood mechanisms by which the

RKN infection begins and expands (Bellafiore et al., 2008; X.-R. Wang et al., 2012; Rutter et al., 2022).

1.2.3 Agricultural impact and management

RKN are responsible for billions of dollars of agricultural crop losses each year (Bernard et al., 2017), and frequently threaten food production and security in the developing world (Sikandar et al., 2020, 2022). They are considered the most damaging PPN group (Jones et al., 2013). Although the most damaging species are generally restricted by ecology to tropical climates, where surface soil temperature rarely falls below 20°C and winter temperatures remain above freezing (Hogger & Bird, 1974; Vrain et al., 1978), there are growing concerns that the worsening global climate crisis could enable range expansion of these species, enabling infection of currently temperate regions where the environment was historically too cold for them to establish populations and thrive (Bebber et al., 2013, 2014; Elad & Pertot, 2014; Dutta & Phani, 2023). This concern makes understanding the biology of the species, particularly its evolutionary origins, methods of adaptation, and suite of host-manipulation effectors vital to future global food crop security.

Many control measures exist for managing RKN infection, although no one measure can decisively tackle a widespread infection, and complete purging of RKN from infected fields and areas has proven to be extremely difficult, as well as time and resource intensive. Most methods of control can be divided into two forms: cultural control and chemical control. Cultural controls include the following: (1) the planting of resistant crops that have either an innate or genetically engineered resistance to RKN infection. This starves the nematode for a season, reducing their population size (Barbary et al., 2015; Al Abadiyah Ralmi et al., 2016). (2) planting of cover crops such as cowpea or marigolds, which the RKN is unable to infect, again reducing the population size (Navarrete et al., 2016; Acharya et al., 2021; Marquez & Hajihassani, 2023). (3) soil solarisation, whereby the soil is covered in transparent plastic film, increasing the soil temperature and killing the eggs of RKN (Rudolph et al., 2023). Chemical controls include the introduction of either fumigant or non-fumigant chemicals to the infected field (Soltani et al., 2013; Hussain et al., 2016), with the former having been observed to be more effective, with the caveat of being phytotoxic and requiring a lapse period between treatment and growth (Noling, 2014). Despite being the most commonly used method of treatment, a drawback of chemical control methods is their environmental impact (Kim et al., 2018; Xiang et al., 2018).

There are several alternative methods of control alongside those outlined above. The first is trap cropping, whereby a RKN susceptible crop is grown in the infected field. RKN infect this crop, and before enough time has passed for mature egg masses to develop, the crop is removed from the field, along with the RKN infecting it. This dramatically cuts the size of the RKN population in the field and stymies the extent of the infection through depopulation and artificial arrest of their reproductive cycle (Westerdahl, 2018). The second is introduction of bacterial or fungal pathogens of RKN to the infected area (Urek, 2013; Forghani & Hajihassani, 2020). These organisms infect or trap the RKN, reducing their ability to infect and damage a host plant, and are considered among the most environmentally safe methods of treatment. The final alternative method is the genetic engineering of plant cultivars to contain genes known to bestow resistance to RKN infection.

Plant cultivars can be engineered to contain genes that bestow resistance to RKN or wider PPN infection, termed resistance genes (*R* genes). These *R* genes tend to operate under a dominant-recessive interaction, bestowing resistance from a single dominant allele (Williamson & Kumar, 2006). *R* genes can arise naturally in populations and be selected for through repetitive exposure to RKN infection, or can be introgressed into a cultivar's genome purposefully with the aim of endowing the introgressed cultivar with resistance to RKN. *R* genes have been identified from many of the host Families, controlling infection from the majority or the most impactful *Meloidogyne* species (Barbary et al., 2015).

Several lineages of RKN have been identified that can evade this genetically engineered resistance and once again infect these crops, raising concerns about such an adaptation arising and being perpetuated in the wild (Eddaoudi et al., 1997; Tzortzakakis et al., 2005; Joseph et al., 2016; Ploeg et al., 2023). An example of this is in the VW4-VW5 system of *M. javanica*, wherein VW4 populations are unable to infect tomatoes holding the *Mi-1 R* gene, whereas populations of VW5 can evade this resistance, albeit with reduced fitness compared to VW4 on non-resistant cultivars (Gleason, 2003; Gleason et al., 2008; Gross & Williamson, 2011). To better develop genetic control methods to combat RKN infection, it is vital that we attain a better understanding of the genomics of RKN and the genetic mechanisms by which they cause infection.

1.3 Genomic analysis

1.3.1 Applications and advancements

Genomics is the study of the complete DNA complement of an organism: that is, the entire DNA sequence as contained in every somatic cell. Genomics, as compared to genetics, is a relatively new field which is rapidly expanding and increasing in complexity and scope year on year (Weissenbach, 2016; McGuire et al., 2020). With genomics, researchers can compare organisms molecularly genome to genome, rather than performing species comparisons based on morphology, phenotype, or individual genes, and has allowed deeper investigation of speciation (Seehausen et al., 2014; Bock et al., 2023), population dynamics and conservation (Allendorf et al., 2010; Formenti et al., 2022), as well as insight into evolutionary origins and mechanisms (Rokas & Abbot, 2009; Reid et al., 2021) and access to genome-wide comparative analysis (Chen et al., 2021; Wright et al., 2024).

Over the last decade genomic technologies have advanced measurably and genomic analysis using resources such as chromosome-scale genome assemblies and high accuracy, high depth, whole genome sequencing has become the norm (Deakin et al., 2019; Mathers et al., 2021; Rabanus-Wallace et al., 2021; Song et al., 2023). Whereas in the past, sequencing technologies were limited to a resolution of hundreds of base pairs, the advent of high-throughput long-read technologies has increased the range of a single sequencing read from hundreds, to tens of thousands of base pairs, reducing the difficulty and increasing the accessibility of investigation of large scale structural variation, as well as greatly increasing the contiguity of assembled genomes (van Dijk et al., 2023). The unprecedented number of chromosome-scale genomes now available is enabling comparative genomic analyses at a scale not seen before. Similarly, the ease in which data concerning the secondary characteristics of the genome, such as base modification and chromatin conformation, can be gathered has made in depth comparative genomics a possibility (Shim et al., 2015; Sati & Cavalli, 2017; Yuen et al., 2021; Stephenson-Gussinye & Furlan-Magaril, 2023). The research methods discussed in this thesis includes genome assembly, genome annotation, and comparative genomic analysis; methylation, synteny analysis, phylogenomics, and testing for selection.

1.3.2 Genome assembly

Genome assembly is the process of creating a textual representation of an organism's genome from the actual molecular genome (Kitts, 2002; Pop, 2009; Li & Durbin, 2024). It is constituted of a series of DNA sequences in a text file; a format called FASTA (Pearson & Lipman, 1988). This enables us to investigate the biology of an organism's genome. With an assembly as a template, sequences and reads from other individuals can be mapped, allowing us to compare multiple genomes like-for-like. There are two main aims of genome assembly: completeness and contiguity (Li & Durbin, 2024).

Contiguity refers to how closely the genome assembly resembles the biological genome in terms of fragmentation. An ideal level of contiguity is that the amount of sequences in the assembly will be congruent with the amount of chromosomes in the genome (Thrash et al., 2020; Wang & Wang, 2022; Walve, 2023). For chromosome-scale comparisons required for analysis of chromosome evolution, a chromosome-scale assembly is typically required, especially if there is no closely related conspecific with which to use as a genomic map. For other analysis, such as analysis of population diversity, evolutionary history, or selection, a lower level of contiguity is required. In contrast, completeness refers to how accurately the assembly compares to the genome in terms of content, rather than structure (Thrash et al., 2020). A genome is considered biologically complete if all content of the genome - genes, transposons, repeats, etc - are contained within the assembly, including telomeres at chromosome termini. This is important for many downstream analyses.

The basic stages of the process of genome assembly are described here, although depending on the biology of the organism, the exact methods can vary widely, and range from wholly automated to including lots of manual curation (Rice & Green, 2019; Thrash et al., 2020; Li & Durbin, 2024). The assembly process begins with DNA being extracted from an organism and sequenced. The sequenced reads are processed and cleaned based on length and quality score (Chen et al., 2018), then fed into an assembly software that combines the reads to generate an assembly. The exact package used can depend on the biology of the genome, the type of sequencing technology used, and the preference of the researcher performing the assembly (Jung et al., 2020; Dida & Yi, 2021).

There are two algorithms that are most commonly used by these packages - overlap layout consensus (OLC) and de Bruijn graph (DBG), however all follow the principle of overlapping reads or k-mers to generate a series of consensus sequences (Kitts, 2002; Pop,

2009). Each of these consensus sequences is referred to as a contig - from contiguous. Collections of contigs make up the assembly. As stated above, ideally there will be a single contig for each chromosome, producing a chromosome-scale assembly. Unfortunately, this is seldom the case and further computational steps are required to increase the contiguity towards that observed in the biological genome, and reduce fragmentation - the antithesis of contiguity (Li & Durbin, 2024).

Several aspects of genome biology can increase the probability of the assembly inaccurately representing the genome. Complicated genomic features and biology can introduce errors, called misassemblies (Meng et al., 2022). Collapse is one type of misassembly, whereby several distinct regions of the genome are represented in the assembly as a single region, often due to high similarity between their representative sequence reads. This is seen most often in areas rich in repetitive sequence, where reads from several identical repetitive stretches littered around the genome are misinterpreted by the assembler as many reads of the same region, and as a result are assembled into only one representation. This is more often seen in assemblies generated with short-reads, as long reads are more likely to span the repetitive region, giving a definitive representation of its true length (Wang et al., 2021).

Fragmentation is another example of misassembly, whereby breaks are introduced into a DNA sequence that would otherwise be contiguous within the organism. This limits the power and accuracy of synteny analysis, examination of large structural changes, and prevents measuring completeness through comparison to cytological chromosome counts. Fragmentation can also lead to the exclusion of intergenic regions, which can contain vital genomic information and limit the quality of annotation. Fragmentation can be reduced through the application of scaffolding and gap closing methods, using long reads or chromatin contact information (Howe et al., 2021; Peona et al., 2021; Whibley et al., 2021). Another consideration to be taken into account during the assembly of a genome is the ploidy of the organism being assembled (Wang et al., 2023). Ploidy refers to the number of copies of a genome contained within the cell, and can vary widely even between closely related species, and sometimes within species (Figure 1.6). Genomes with greater than two copies (diploid) are said to be polyploid, of which there are two general classifications: autoploidy and alloploidy. Autoploidy is where the duplicated copies share a homologous origin and are highly similar, usually resulting from a whole genome duplication (WGD) event within the individual – potentially generations ago - or immediate parent. This WGD can arise from incomplete chromosome segregation post replication or a disruption in the cell

cycle preventing anaphase, but the end result is unreduced gametes. These gametes fuse, generating an interspecific polyploid offspring

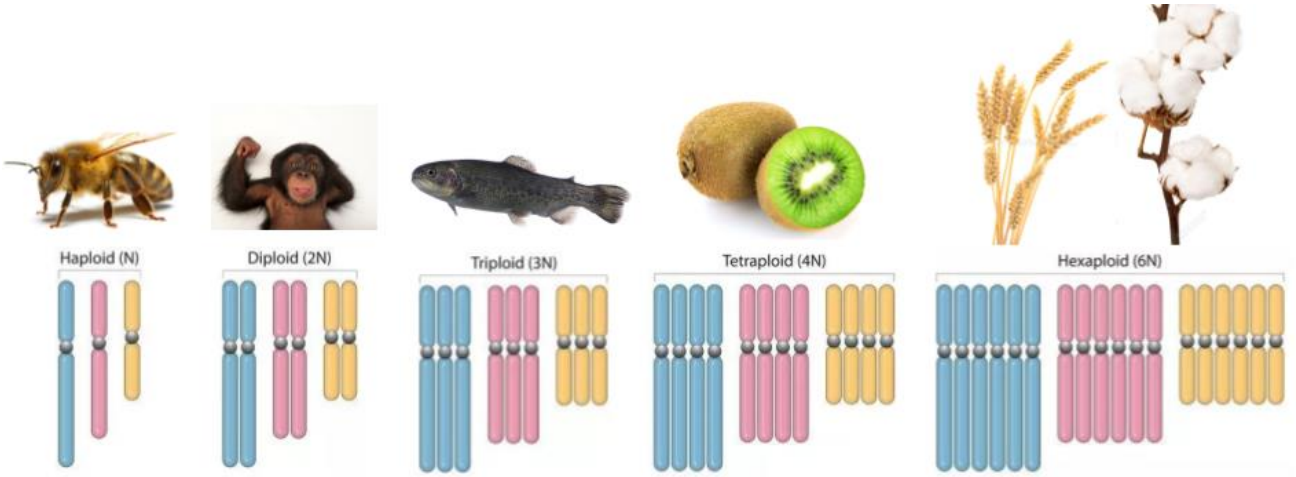


Figure 1.6: Different ploidy within organisms. Ploidy states vary between organisms. In this example, the haploid complement is three chromosomes (2n = 1x = 3). As ploidy increases, so does the chromosome number, although additional chromosomes are not distinctly different, but additional copies to the haploid complement. Eg, diploid is two copies (2n = 2x = 6), triploid is three copies (2n = 3x = 9), tetraploid is four copies (2n = 4x = 12), etc.

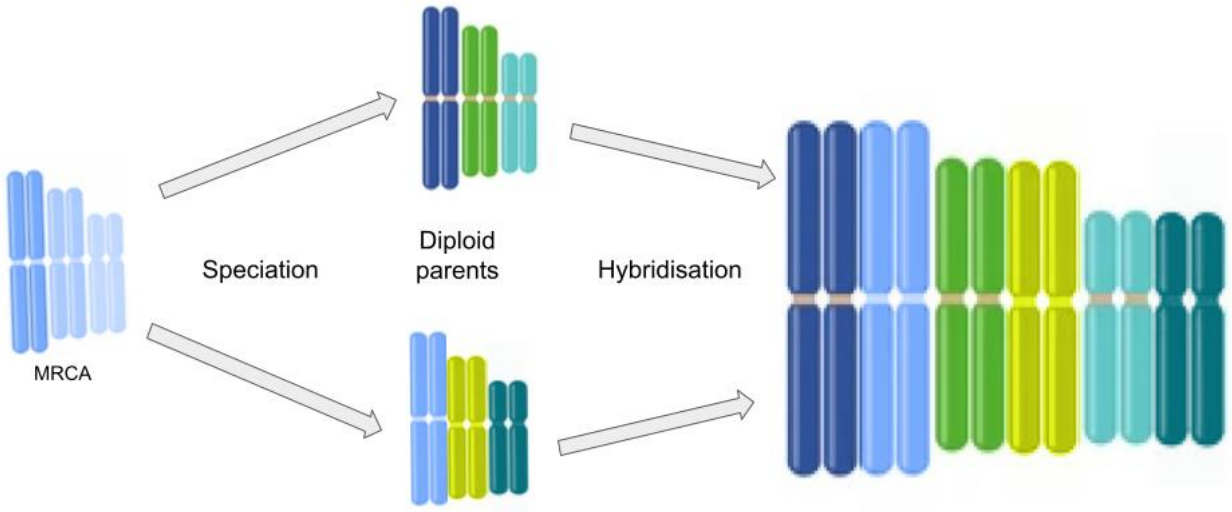


Figure 1.7: Diagram of allopolyploid formation. The most recent common ancestor (MRCA) of the extant allopolyploid splits into two distinct species. These novel species - the parent species of the allopolyploid - diverge in isolation, evolving distinct genome sequences and architecture. The parent species then hybridise, producing allopolyploid offspring. Chromosome groups obtained from each parent species are referred to as subgenomes, with each subgenome representing the genome of each parent contributing to the hybridisation.

with low heterozygosity and genetic diversity (Bretagnolle & Thompson, 1995; Lv et al., 2024). Allopolyploidy is where the multiple copies are the result of hybridisation of two or more closely related species, and share less sequence similarity than autopolyploids. Allopolyploids are said to have subgenomes, one from each parent, and each far less similar to the other than would be seen in an autopolyploid (Figure 1.7; Stebbins, 1947; Frawley & Orr-Weaver, 2015).

Complex ploidy can lead to misassembly (Whibley et al., 2021; Kong et al., 2023). Multiple diverse copies at many loci can confuse the assembly software and introduce what are known as "phase switch errors"; incidences within the assembly where the sequence switches from representing one copy of the genome to another. Switch errors are impossible in an assembly of a haploid organism, but become more likely as the ploidy of the organism increases or as sequence divergence decreases. In polyploid organisms switch errors can become more prolific, although this can be managed and prevented through the application of long read technologies and their resultant reads. As long reads overlap, the chance of a switch error becomes less likely due to phased blocks becoming less fragmented. A single long read is essentially a phased block of the genome. Combined with the increased capacity for overlap with other long reads, this reduces the opportunity for haplotype switching across an assembled contig (Wenger et al., 2019; Duan et al., 2021). This is further complicated in allopolyploids where homeologous recombination - the exchange of sequence information between subgenomes - is possible (Mason & Wendel, 2020).

Another phenomenon that can impact the quality of a genome assembly is contamination, wherein the assembly contains sequences from a non-target, undesired species (Merchant et al., 2014; Kryukov & Imanishi, 2016). Contamination of a genome assembly can arise through several pathways, one of which is contamination of target DNA during extraction or sequencing from organisms that naturally reside within or alongside the target species. For example, when extracting DNA from very small organisms it is common to also extract the DNA of microorganisms living inside or upon it, such as bacteria or other parasites (Kumar & Blaxter, 2011). This contamination from symbionts can happen inversely, where DNA of the host or immediate environment of the target species is sequenced alongside that of the target. This is common the case of root-knot nematodes, where many studies report *Solanum lycopersicum* in their draft assemblies due to it being the host on which many RKN are cultured (Blanc-Mathieu et al., 2017; Szitenberg et al., 2017; Susič et al., 2020; Winter, 2020; Winter et al., 2024). Contamination can also be introduced to a sequencing library from organisms present within the lab during the

extraction and sequencing process, or from the researcher themselves (Salter et al., 2014; Cornet & Baurain, 2022).

Contaminants within a read library or assembly can be identified through alignment against a database of sequences with known taxonomy, and assigned a taxon based on sequence similarity, and removed. There are several open access, publicly available methods to perform this (Laetsch & Blaxter, 2017; Challis et al., 2019; Lu et al., 2022). However, another consideration when screening for contamination in a genome assembly is the accuracy and robustness of the databases used to assign taxonomy to a read or contig. These databases can contain misclassifications, which can then propagate into contamination screening, leading to sequences being erroneously removed from the assembly (Lupo et al., 2021). Another reason taxonomy of a contig can be misassigned, although far less common, is when regions of a non-target species have been inserted within the genome through horizontal gene transfer (Trappe et al., 2016; Brito, 2021). This seemingly occurs frequently in RKN, where many studies report a considerable amount of sequences assigned as Arthropoda by taxonomic databases, in much larger amounts than could realistically be contamination. Alongside this, these Arthropoda-assigned sequences share the same guanine-cytosine ratio and coverage depth of the target species, leading to the suspicion that these assignments are taxonomic errors (Szitenberg et al., 2017; Koutsovoulos et al., 2020; Susič et al., 2020; Winter, 2020; Somvanshi et al., 2021; Mota et al., 2024; Winter et al., 2024). Careful assessment of the likelihood of species identified as contaminants can help identify unlikely or suspicious assignments, but despite this taxonomic errors complicate the process of identifying and removing contaminants. Moreso, contaminants that are not detected are typically then uploaded to the same taxonomic databases used to detect the errors, propagating the issue (Lupo et al., 2021; Cornet & Baurain, 2022).

1.3.3 Genome annotation

Annotation is the process of detecting and recording of discrete genomic features, such as genes, motifs, repeat elements, telomeres, and more. There are two facets of gene annotation: structural and functional. Structural annotation of genes and motifs is performed computationally using either machine learning techniques, mapping of sequenced features, or a combination of both. There are pipelines available to automate and combine these methods, the most widely used being *MAKER3 and BRAKER3* (Campbell et al., 2014; Gabriel et al., 2023). Structural gene annotation by machine learning is termed *ab initio*

annotation and is achieved through identification of regions sharing similarity to known genes of model species, as well as similarity of features with known genic motifs. Its primary benefit is that it can be used in the absence of any other information about the genes or proteins of the species. Although a useful option to have in this absence, the more divergent or evolutionarily distant the target species is from the reference genes of the model, the weaker and more inaccurate the annotations become. Examples of *ab initio* annotation software are *GeneMark, Exonerate,* and *Augustus* (Besemer & Borodovsky, 2005; Slater & Birney, 2005; Hoff & Stanke, 2019).

A more accurate result can be attained through combination of *ab initio* methods with sequencing information from the genes of the target species. RNA-seq - the sequencing of mRNA transcripts from the organism - can facilitate the generation of a library (transcriptome) of the organism, for which the genomic positions can then be determined with mapping or alignment approaches. With short read technologies these reads require assembly into whole transcripts, which has the potential to introduce errors and lower the quality of final annotation (Hölzer & Marz, 2019). Recent advances in long read sequencing, resulting in the generation of PacBio Iso-Seq sequencing, has enabled the sequencing of whole transcripts, from guanine cap to poly-A tail, removing the need for assembly and increasing the accuracy of the transcriptome and subsequent genome annotations (Beiki et al., 2019; Ali et al., 2021).

Following this, structural annotations need to be assigned ontology, which is performed through functional annotation (Yandell & Ence, 2012). This is the process of assigning function to the transcriptome. This is important primarily for linking structural changes within or between genomes to a predicted change in phenotype. There are many approaches to functional annotation with a range of accuracy and computational difficulty (Kiritchenko et al., 2005; Sivashankari & Shanmughavel, 2006; Zdobnov et al., 2021).

Another type of feature requiring annotation are repeats. Repeats are regions of the genome containing stretches of DNA with a repeating sequence. These repeating sequences can occur in tandem - such as short and long tandem repeats (STR and LTRs, respectively) - and make up long stretches of the genome sequence, or can be transposable repeats, where multiple copies are found across the genome. As with genome assembly, repeats can cause issues with annotation, particularly *ab initio* annotation. As such, an important stage of preparing the assembly for gene annotation is repeat annotation (Kitts, 2002; Jiang, 2013). Regions identified as repeats are "masked"; they are changed to

lowercase in the FASTA file of the genome assembly. This masking is detected by gene annotation pipelines and the masked regions are excluded from the process due to the extra computational requirements required for detecting genes in regions that are likely devoid of them. The most widely used software for repeat annotation is *RepeatMasker*, which detects repeats through alignment of a repeat database to the genome assembly. As with *ab initio* gene annotation, the accuracy of this method is dependent on similarity of repeats to those in the database (Tarailo-Graovac & Chen, 2009).

1.3.4 Comparative genomics

Comparison of whole genomes can give us a detailed view of how species and populations differ at both a genetic and genomic level. Comparative genomics is the process of comparing two or more genomes and performing a suite of exploratory genomic analyses to detect differences. These differences can include, but are not limited to, basic descriptive statistics such as genome size, number of genes, and chromosome number, as well as ploidy, selection pressures, structural or genic conservation, codon usage frequency, base usage and guanine-cytosine ratio, differences in epigenetics, and through sequence comparison of many genes, phylogenomics and inference of the evolutionary history of a group (Sommer & Streit, 2011; Li et al., 2016; Pratx et al., 2018; Bybee et al., 2021; Wright et al., 2024). The power of comparative genomics has increased massively in line with improvements to the length and quality of sequencing reads - both DNA, RNA, and protein - and subsequent increases in accuracy, completeness, and contiguity of genome assemblies and annotations (Kille et al., 2022; Miga & Eichler, 2023; Li & Durbin, 2024).

1.4 The *Meloidogyne incognita* group

The *Meloidogyne incognita* group (MIG) is a group of species in clade I of *Meloidogyne* known for their remarkable agricultural impact, even compared to other PPN. The MIG includes the groups namesake *Meloidogyne incognita*, as well as *Meloidogyne arenaria* and *Meloidogyne javanica,* all of which are positioned phylogenetically within clade I of *Meloidogyne* (Figure 1.2). Found on all continents with mild winter temperatures (Trudgill & Blok, 2001), these species are all apomictic asexuals, reproducing through mitotic division (Triantaphyllou, 1963, 1981; Janssen et al., 2016).

1.4.1 Diversity of genomic architecture in *Meloidogyne* and the MIG

Species in the *Meloidogyne* genus exhibit a wide range of genomic constitutions ranging from diploid to polyploid (Eisenback & Triantaphyllou, 1991). Some species have diploid genomes, and undergo meiosis and recombination (Goldstein & Triantaphyllou, 1982; Janssen et al., 2017; Álvarez-Ortega et al., 2019). Other species within the genus have very complex genomic arrangements, of which the MIG stand out for their genomic complexity despite their frequent study (Lunt, 2008; Castagnone-Sereno et al., 2013; Lunt et al., 2014; Blanc-Mathieu et al., 2017; Szitenberg et al., 2017).

The MIG contains several species that appear to be the descendants of a past hybridisation event between two previously existing diploid species (Lunt, 2008; Lunt et al., 2014; Szitenberg et al., 2017). This has left these species with divergent homoeologous subgenomes within one organism, resulting in many members of the MIG containing multiple, often redundant copies of many genes (Bird et al., 2018; de Tomás & Vicient, 2023). These subgenomes are suspected to be more closely related to the corresponding homologous subgenome of other species of the group than they are with each other, as a result of their shared hybrid origins (Szitenberg et al., 2017).

1.4.2 Sexual systems

Another factor increasing the complexity of some MIG genomes is their sexual system. Many clade I species - including *M. arenaria, M. incognita,* and *M. javanica* - are apomicts and reproduce asexually, producing clonal offspring (Castagnone-Sereno & Danchin, 2014). Apomixis is a form of reproduction whereby offspring are formed through mitotic division of a germline cell, in a situation of "reproduction without fertilisation" (Lincoln et al., 1983). This leads to offspring being genetically identical to the parent. An exception to this in the MIG is *M. floridensis,* which is automictic (Handoo et al., 2004; Jaron et al., 2018), however some populations are believed to be apomictic (Jaron et al. 2018). Automictic reproduction is also parthenogenetic, but in contrast to apomixis the offspring undergoes meiosis, allowing recombination to take place (Mogie, 1986). This produces offspring that contain the same genomic content as the parent, but with the possibility of recombination facilitating crossover of regions between copies.

1.4.3 Genomic resources

Many clade I *Meloidogyne* assemblies and their subsequent analyses are outdated compared to their contemporaries generated since advances in long-read sequencing technology and other bioinformatic methods. This is particularly true of the apomictic polyploid MIG species, whose genomic complexity have made genome assembly and interpretation more difficult.

Alongside this, no standard has been set with regard to representation of the ploidy of the genome, with many MIG assemblies prior to 2020 being agnostic of ploidy or copy number. Not considering how the assembly represents ploidy and intragenomic diversity is a widespread problem across the reporting and announcing of genome assemblies (Li, 2021), made more observable with the recent ability to properly phase polyploid assemblies (Kong et al., 2023; Li & Durbin, 2024). In the era of short-read sequencing, assemblies were typically published as the haploid version of the genome: a haplome. This is accurate for haploid species, or diploid species with low heterozygosity, or assemblies where multiple copies are similar enough that inferring heterozygous regions requires only base calling. With polyploids however, genomic diversity is often too great to denote a single haplome, and the lack of definition of this in genome announcements makes the actual content of some assemblies unclear. New terminology coined by Li (2020) introduces terms able to properly describe the type and content of assembly, including dual assembly, haplotype-resolved assembly, etc.

This is particularly an issue for the MIG, where comparative genomics has been performed in an unintentionally abstract way, between species and assemblies that may not actually be comparable and with an incomplete awareness of subgenome specific characteristics. Assemblies generated from short read sequencing of an allopolyploid would be extremely prone to phase switch errors, as well as collapse where assembly algorithms applied are not designed to handle and separate polyploid haplotypes. Combined with uncertainty as to which subgenomic haplotypes the assembly represents, nor which contigs belong to which subgenome, accurate comparative genomics is difficult.

Since the completion of the research described in this thesis a paper was published claiming to present phased and chromosome-scale genome assemblies of *M. arenaria, M. incognita,* and *M. javanica* (Dai et al. 2023). Alongside this a comparative genomic analysis was performed similar to those performed in Chapter 3 of this thesis. The research performed in this thesis' data chapters was completed before publication of the Dai et al

(2023) paper, and as such the results and findings within them have not been compared to this work. Our findings are however addressed with the context of Dai et al. (2023) in the general discussion chapter (Chapter 5).

1.5 Reproducibility

Reproducibility refers to the spectrum of potential for a study or method to be repeated. Over recent years the concept of tangible reproducibility has been pushed to the forefront of STEM research. This is partly due to the "crisis of reproducibility" experienced by the social sciences, and emerging in other fields (Hudson, 2021).

The FAIR principles are a list of tenets established as a framework for reproducible research. These principles are as follows: F, findable. It should be possible to find or access methods, results, data, and metadata, in a format that is somewhat machine-readable. A, accessible. The aforementioned aspects of the study should be accessible and not withheld behind paywalls or other authentication systems, ie, data and metadata should be publicly accessible. I, interoperable. Data and metadata should be produced, assessed, and stored in industry standard formats (FASTA, CSV, GFF, etc) to make reanalysis and replication attainable. R, reusable. Data and metadata should be well-described to facilitate reuse (Wilkinson et al., 2016).

One method of ensuring reproducibility in bioinformatic analysis is by using automated workflows and pipelines such as *snakemake* and *Nextflow* (Köster & Rahmann, 2012a; Di Tommaso et al., 2017). Doing so reduces the chance and impact of human error, as well as streamlining initial and future analysis (Spjuth et al., 2015; Wratten et al., 2021). Parameters of modules in the workflow can be controlled from a top level configuration file. This makes performing slightly altered iterations accessible, controllable, and easily recordable. In this work we apply these techniques wherever possible, and produce several automated bioinformatic workflows to perform our analyses.

In some analysis the development and use of an automated workflow is either not necessary or unachievable, often due to aspects of a process requiring heuristic interpretation, or from an inability of a software to be nested in a workflow. In such cases the best recourse is to record the bioinformatic process in an electronic notebook. Throughout the work presented here, where automated workflows have not been attainable we have performed analysis in Jupyter notebooks (Kluyver et al., 2016). By doing so, all

code used in analysis is recorded and easily reproducible when provided with the appropriate data and metadata. In some cases, often where a part of analysis was not suitable to a workflow or notebook, we have written unique standalone scripts.

An important and sometimes overlooked aspect of reproducible bioinformatic analysis is version control: recording and maintaining the softwares used, along with their versions (Kulkarni et al., 2018). As new versions of software are developed, the finer functions of the software may change, and this can introduce differences to analytical outputs, especially when attempting to reproduce an analysis that was completed in the past. To manage this we have utilised *conda* throughout our research (Pflüger, 2019). *Conda* generates computational environments where each software installed is version controlled, with the software and version recorded in metadata. Alongside assisting in managing incongruent dependencies between software and ensuring functionality, using *conda* environments ensures that each time an analysis is run it uses the exact same software and underlying code as previous analyses. Another beneficial aspect of working in this way is that the *conda* environment metadata can be exported alongside the data and metadata of the entire analysis, enabling accurate reproducibility in the future.

Data backup and storage is another important aspect of reproducibility and is especially important when considering adherence to FAIR principles (Wilkinson et al., 2016; Peng & Hicks, 2021; Katz et al., 2022). Aside from insuring against accidental data loss or damage, uploading and storing data in publicly accessible, free-use databases facilitates reproducibility and ensures the data is maintained and available long-term. The strongest examples of this are NCBI GenBank (Sayers et al., 2020) and the Sequence Read Archive (Katz et al., 2022), without which some of the research in this thesis would not have been possible, and to which all genomic resources generated within this work are uploaded.

1.6 Research questions

The quality and accuracy of historic genome assemblies of MIG species has been limited by the ability of the available technology to handle the complexity of their genomes. Recent advancements in long-read sequencing and chromatin capture methods have revolutionised genome assembly, leading to the frequent production of chromosome-scale assemblies, including those of allopolyploids (Mitros et al., 2019; Cerca et al., 2022; Kuhl et

al., 2022; Li & Durbin, 2024). In this study, we aim to improve upon past efforts to assemble the genome of *Meloidogyne javanica* by leveraging these long-read technologies and modern bioinformatic methods.

A key requirement for a proper genomic investigation of an allopolyploid species is the identification and phasing of its subgenomes. With current methodologies, it is possible to phase allopolyploid assemblies and assign contigs and scaffolds to their respective subgenomes, enabling accurate subgenomic comparisons. In this research, we will attempt to apply these methods to the genome assembly of *M. javanica* and utilise the newly phased assembly to detect and analyse differences between subgenomes.

Methylation has been suggested as an important mechanism through which a novel polyploid manages genomic shock and increased gene dosage (de Tomás & Vicient, 2023). However, conclusive evidence of methylation in *Meloidogyne* species is lacking (Pratx et al., 2018). By using methylation-aware base-calling developed by Oxford Nanopore Technologies, we will attempt to detect and explore the presence of methylation in *M. javanica.*

Further, subgenomes of the MIG have been suggested to be more similar at a nucleotide level to their homologs in other species than to their homeologs within the same species (Szitenberg et al., 2017; Winter, 2020). Based on this observation, we hypothesise that it should be possible to align other MIG genome assemblies to a phased assembly and assign phase to contigs of these other assemblies. If successful, this approach will enable more detailed and confident comparative genomic analyses of subgenomes of the MIG than attainable before. We will test this hypothesis and use the resulting phased contigs sets to assess similarities or differences between MIG species overall and between subgenomes, investigating subgenome synteny across species and codon usage frequency divergence.

Subgenomic dominance, often observed in allopolyploids, serves as a mechanism to regulate gene dosage and maintain functional pathways. This dominance has been shown to facilitate copy loss and diploidisation, and different pathways to subgenome dominance could explain the variance in ploidy predicted across the MIG species (Mandáková et al., 2016; Li et al., 2021). One way in which subgenome dominance can vary is through different selection pressures across species, leading to a difference in which genes are under selection on either subgenome (Bird, 2022). Using tests of selection, we attempt to identify these differences, and compare if they are congruent across species. We also attempt to

detect the signs of subgenome dominance on a gene scale, where it would be expected that as one copy becomes dominant, the alternative copies are susceptible to gene fractionation.

The *Cg-1* region of *M. javanica* has been identified as involved in gain-of-virulence mutations in the VW5 lineage (Gleason et al., 2008; Gross & Williamson, 2011), though the relevance of the *Cg-1* region and mechanism of action is unclear. Utilising advanced sequencing technologies and bioinformatic approaches, we will attempt to describe the *Cg-1* region in greater detail than previously possible. Additionally, we explore the potential for other genomic differences between VW4 and VW5 that were undetectable using older technologies and resources, and ascertain if any potential differences could explain the observed difference in virulence.

1.7 Description of chapters

Chapter 1: General introduction

This chapter introduces and describes all relevant topics required for understanding and discussion of the thesis. We introduce plant parasitic nematodes and their biology, with a comprehensive summary of past research into their life and evolutionary history, narrowing onto the *Meloidogyne* genus; root knot nematodes (RKNs). Genomics is then introduced and discussed, including the foundational concepts of genome assembly, annotation, and comparative analysis. We then focus on specific aspects relating to nematode genomic research; the lack of resources and the complexity of some MIG genomes. Reproducibility is then introduced with specific attention brought to reduction of difficulty when analysing complex groups such as *Meloidogyne*, and the impact provision of reproducible tools can have on the field. We discuss how the lack of tools and difficulty of use has stymied growth in this field. Finally, the larger research questions of the thesis are introduced.

Chapter 2: Phased chromosome-scale genome assembly of an asexual allopolyploid root-knot nematode reveals complex subgenomic structure

Chapter 2 lays out the process of generating a chromosome-scale assembly of the MIG RKN *Meloidogyne javanica*. In it, we describe the methods used and the results of our assembly and annotation of the genome, as well as providing comparative statistics to other

contemporary *Meloidogyne* genome assemblies. We describe how we phased scaffolds of the genome assembly into subgenomes through the detection of ancestral repetitive k-mer sequences. We go on to perform a synteny analysis between the subgenomes in order to determine the amount of gene conservation and collinearity remains between the homoeologs. Our overall outputs are a phased chromosome-scale genome assembly of *M. javanica* and corresponding gene annotations, alongside discovery that large amounts of synteny and collinearity have been conserved between the subgenomes.

Chapter 3: A comparative analysis of subgenomes in allopolyploid root-knot nematodes of the *Meloidogyne* genus

In this chapter we perform several comparative analyses between subgenomes and species of the MIG and *Meloidogyne luci*, including ploidy analysis, differences in selection pressures, and preservation of synteny between subgenomes of these species. We attempt to infer which subgenome is dominant, experiencing a lower level of gene fractionation compared to the other and performing the majority of genomic processes. We also perform additional comparisons between the subgenomes of *M. javanica*, observing methylation and detecting paired homoeologous genes under differential selection pressures. Our findings include the definitive ploidy state of these species and through tests of selection we find a balanced state of subgenome dominance arising from oppositional substitution rates at transcriptional loci. We place *M. luci* in a wide phylogenomic framework for the first time and find that both of its subgenomes appear to be homeologous to the subgenomes of the MIG, strongly suggesting a shared origin with the MIG. We also determine that subgenomes across species retain large blocks of synteny, as well as detecting signals of methylation in *Meloidogyne javanica.*

Chapter 4: Identifying candidates for gain-of-virulence mutations in *Meloidogyne javanica*

In the fourth chapter we investigate a lineage of *M. javanica* known to be able to evade genetically engineered resistance in tomatoes, the VW5 lineage. Beginning by characterising a region thought to be involved in resistance evasion by previous published work, we go on to identify a large deletion on one chromosome, encompassing this previously identified region as well as dozens of other potential candidate genes, some of which appear to be completely deleted from VW5. We characterise this deletion and the

genes within it in order to provide a foundation for further research to investigate these newly discovered and highly promising candidates of gain-of-virulence mutation.

Chapter 5: General discussion

In the final chapter we collate our findings and interpret them in light of one another. We present contributions to the field produced by this work and their place in the wider literature, as well as discussing and suggesting future directions for research. In this chapter we also discuss novel concepts suggested by recent publications that are supported by incidental findings of this work, and address our efforts to work reproducibly in an open framework.

Chapter 2: Phased chromosome-scale genome assembly of an asexual, allopolyploid root-knot nematode reveals complex subgenomic structure

Michael R Winter[1*], Adam P Taranto[2], Henok Zemene Yimer[3], Alison Coomer Blundell[2], Shahid Siddique[3], Valerie M Williamson[2], David H Lunt[1]

[1]School of Natural Sciences, University of Hull, Hull, UK

[2]Department of Plant Pathology, University of California Davis, Davis CA, USA

[3]Department of Entomology and Nematology, University of California Davis, Davis CA, USA

Correspondence: mrmrwinter@gmail.com (MRW)

2.0 Abstract

We present the chromosome-scale genome assembly of the allopolyploid root-knot nematode *Meloidogyne javanica*. We show that the *M. javanica* genome is predominantly allotetraploid, comprising two subgenomes, A and B, that most likely originated from hybridisation of two ancestral parental species. The assembly was annotated using full-length non-chimeric transcripts, comparison to reference databases, and ab initio prediction techniques, and the subgenomes were phased using ancestral k-mer spectral analysis. Subgenome B appears to show fission of chromosomal contigs, and while there is substantial synteny between subgenomes, we also identified regions lacking synteny that may have diverged in the ancestral genomes prior to or following hybridisation. This annotated and phased genome assembly forms a significant resource for understanding the origins and genetics of these globally important plant pathogens.

2.1 Introduction

2.1.1 The assembly of allopolyploid genomes

Allopolyploidy is a genomic state characterised by more than two chromosomal complements, with one or more of these complements resulting from a hybridisation event leading to the presence of distinct (homoeologous) subgenomes within a single cell (Glover et al., 2016). Allopolyploids may account for 11% of plant species including many model species and important crops (Barker et al., 2015). Although not as frequent as in plants, genomic investigations are indicating that ancestral genome duplication, hybridisation, and complex genome arrangements are more widespread than previously recognized in animals (Schoenfelder & Fox, 2015; Session et al., 2016). Assembly and analysis of allopolyploid genomes, however, is challenging for a number of reasons (Ming & Man Wai, 2015). The increased number of alleles within an allopolyploid genome can interfere with algorithms used by many assemblers leading to the accumulation of switch errors; regions of the assembly where the sequence switches between haplotypes or homoeologs. In addition, the high amount of repeat content often found in allopolyploid genomes can result in fragmentation of the final assembly if sequenced reads fail to span the repeat (Kyriakidou et al., 2018; Rhie et al., 2021). Another difficulty in accurate assembly of allopolyploid genomes has been 'phasing' i.e., the assignment of assembly contigs to the correct subgenome. Switch errors and misassemblies introduced during the assembly process can impair the signals required to successfully phase a scaffold, and potential crossover interactions between homoeologs can further complicate this signal (Zhang et al., 2020; Saada et al., 2022).

Assembly of allopolyploid genomes has become more feasible due to the advent of long-read sequencing technologies and better assembly algorithms. Most chromosome-scale allopolyploid assemblies in the literature are of agricultural plants (Edger et al., 2019; Gan et al., 2021; Kolesnikova et al., 2022; Zheng et al., 2022), although a few chromosome-scale allopolyploid assemblies of animal genomes are now also available (Du et al., 2020; Kuhl et al., 2022).

2.1.2 Root-knot nematodes

Root-knot nematodes (RKN) - genus *Meloidogyne* - are a group of obligate plant parasites that include species which severely reduce crop yield (Perry et al., 2009). Second-

stage juveniles (J2s) of RKNs hatch in the soil and are non-feeding, needing to invade a host plant root to complete their life cycle. Upon reaching the vascular cylinder, J2s induce the formation of a feeding site inside the root, characterised by formation of a gall ("root-knot") and highly modified "giant cells" on which the nematode feeds (Williamson & Gleason, 2003). Three closely related species within the *Meloidogyne* genus, *M. arenaria*, *M. incognita*, and *M. javanica*, which we refer to here as the *Meloidogyne incognita* group (MIG), have extremely broad host ranges spanning the majority of flowering plants (Trudgill & Blok, 2001; Szitenberg et al., 2017) and together are estimated to cost the agricultural industry tens of billions of US dollars a year (Wesemael et al., 2011; Jones et al., 2013; Gregory C. Bernard et al., 2017). Almost 100 RKN species have been described (Eisenback & Triantaphyllou, 1991; Subbotin, et al., 2021), which differ in host range, pathogenicity, geographic range, morphology and reproductive mode.

Although RKN species have diverse modes of reproduction including amphimixis, automixis, and obligate apomixis, cytological examination indicates that *M. javanica* and most other MIG species reproduce by mitotic parthenogenesis (Triantaphyllou, 1985; Bird, et al., 2009); that is, maturation of oocytes consists of a single mitotic division in which chromosomes remain univalent at metaphase. Phylogenomic analysis has revealed that each species possesses two divergent copies of many genes, that the three species likely originated from interspecific hybridisation, and that they share the same ancestors who have provided the A and B subgenomes (Lunt, 2008; Blanc-Mathieu et al., 2017; Szitenberg et al., 2017).

Despite asexual reproduction, field isolates and greenhouse selections of MIG species are diverse and successful, differing in their ability to reproduce on specific crop species and varieties (Hartman & Sasser, 1985; Roberts & Thomason, 1986; Rammah & Hirschmann, 1990; Wesemael et al., 2011). A widely investigated example is acquisition of ability to reproduce on tomato with the resistance gene *Mi-1*, which confers effective resistance to MIG species and is widely deployed for nematode management in tomato (Williamson & Kumar, 2006). Many independent studies have identified MIG populations that are able to break Mi-mediated resistance; these include both field isolates and greenhouse selections of isofemale lines (Gleason et al., 2008; Hajihassani et al., 2022). However, efforts to decipher the genetic mechanisms for these phenotypic variants have not so far been successful due in part to the lack of tractable genetics and limitations in genome assemblies. Current MIG genome assemblies are fragmented and the homoeologous subgenomes are

mostly unphased (Blanc-Mathieu et al., 2017; Szitenberg et al., 2017; Sato et al., 2018; Susič et al., 2020) making it difficult to compare homoeologous sequences, gain a true picture of diversity, or to understand the nature of functional variation.

Here we apply a combination of modern genomic and bioinformatic approaches to generate a highly contiguous, chromosome-level assembly of *M. javanica*, phased into two subgenomes, creating the first chromosome-scale genome assembly of an apomictic allopolyploid animal that we are aware of. This assembly should provide a very valuable framework for research into the diversity and functional divergence of plant pathogenic nematode species. In the wider research landscape, genomes such as those from within the MIG can aid in our understanding of adaptation, ploidy, and evolution of genomes following hybridisation events and loss of meiosis (Fox et al., 2020).

2.2 Methods

2.2.1 Reproducibility

Wherever possible, this study attempted to contain all bioinformatic processes in reproducible workflows or scripts, for the purpose of openness and enabling replication. Workflows and code are archived in a Zenodo repository along with final outputs (doi: 10.5281/zenodo.10784780). The raw reads plus final nuclear and mitochondrial assemblies are available on the International Nucleotide Sequence Databases (INSDC) (ACC: PRJNA939015) (ACC: GCA_034785575.1).

2.2.2 Biological material

Meloidogyne javanica strain VW4 was used for this work (Gleason et al., 2008; Szitenberg et al., 2017). Cultures of this strain have been maintained on tomato plants under greenhouse conditions for over 30 years (Yaghoobi et al., 1995). Periodic transfers of single egg masses have been carried out to maintain uniformity. For DNA preparation, eggs were harvested from roots and cleaned by sucrose flotation as previously described (Branch et al, 2004) then flash-frozen in liquid N2. High molecular weight DNA (HMW DNA) isolation was carried out at UC Davis Genome Center (Supplementary Methods 2.1). Integrity of the

HMW gDNA was verified on a Femto Pulse system (Agilent Technologies, CA) where majority of the DNA was found to be in fragments above 100 Kb.

Total RNA was isolated from three *M. javanica* life stages: eggs, freshly hatched juveniles, and females dissected from tomato roots 21 days after infection. This material was flash-frozen and RNA was extracted using an Rneasy Kit (Qiagen, USA) following the manufacturer's instructions. TURBO Dnase treatment was carried out to remove genomic DNA from total RNA samples (TURBO DNA-free Kit™, Ambion, USA). RNA concentration and purity were measured using a NanoDrop OneC Microvolume UV-Vis Spectrophotometer (Thermo Scientific, USA). RNA integrity and quality was assessed using a 2100 Bioanalyzer Agilent Technologies G2939BA.

2.2.3 Sequencing and QC

High fidelity (HiFi) long-read sequencing

PacBio HiFi library preparation and sequencing of HMW DNA was performed by UC Davis DNA Technologies Core on a PacBio Sequel II (Supplementary Methods 2.1). Data from the two generated libraries were pooled and only reads longer than 5000 bp and with a quality score over 15 were retained.

Nanopore sequencing

Nanopore sequencing of HMW DNA using Oxford Nanopore Technologies (ONT) systems was carried out by UC Davis DNA Technologies Core using PromethION. The super-long-read DNA sequencing protocol (Supplementary Methods 2.1) yielded 23 Gbp of data, which was then filtered to contain only reads longer than 25 Kbp.

Hi-C chromatin conformation capture

DNA was prepared using Proximo Hi-C kit (Animal) as recommended by the manufacturer (Phase Genomics, Seattle, WA, USA). Library preparation and sequencing were carried out at the UC Davis DNA Technologies Core and scaffolding with Proximo was performed by Phase Genomics (Supplementary Methods 2.1).

PacBio Iso-Seq

Reads from one Sequel II SMRT cell (Supplementary Methods 2.1) were quality controlled and converted into clustered reads using the *IsoSeq3* pipeline with default parameters (PacificBiosciences, 2022).

2.2.4 Genome profiling

Profiling was performed by *Genomescope2* and *smudgeplot* (Ranallo-Benavidez et al., 2020). All quality controlled PacBio HiFi genomic DNA libraries were combined and used as input for these programs. Both were run with default parameters, aside from *-ploidy* in *Genomescope2* set to 4, as this was the primary ploidy indicated by visual analysis of *smudgeplot* and mapped reads.

2.2.5 Assembly

Draft assembly

Initial draft assemblies were generated using several different assemblers and appraised with *asmapp* (Winter, 2022). The most contiguous assembly was carried forward, after which haplotigs – contigs with a homologous allele (A1 and A2, or B1 and B2) -were identified and removed using *purge_dups* with default settings, to leave only one allele per subgenome (Guan et al., 2020).

Scaffolding with Oxford Nanopore and Hi-C

We used *SLR* (Luo et al., 2019) to scaffold the assembly with our trimmed and concatenated ONT reads. The assembly was then scaffolded with Hi-C reads using the *Proximo* pipeline by Phase Genomics Ltd. The resulting contact map was manually curated using *Juicebox* assembly tools (Durand et al., 2016) based on contact linkage information and TAD presence or absence. A third phase of scaffolding was performed by *samba,* to break potential misassemblies introduced during manual *Juicebox* scaffolding and scaffold sequences that were broken as debris (Zimin & Salzberg, 2022).

2.2.6 Annotation

Repeat annotation

RepeatModeler was applied to the assembly (Smit et al., 2015b), with the resulting library of repeat models used as input for *RepeatMasker* to annotate repeat regions and generate a soft-masked version of the genome assembly (Smit et al., 2015a).

Gene annotation

Iso-Seq reads were mapped to the assembly and collapsed using a *snakemake* workflow automating the *IsoSeq3* pipeline (PacificBiosciences, 2022). The annotation pipeline *MAKER3* was then employed to perform *ab initio* and predictive annotation of our assembly (Holt & Yandell, 2011; Campbell et al., 2014). Full methods of annotation iterations and parameters can be found in Supplementary Methods 2.1.

2.2.7 Subgenome phasing

Identification of homoeologous scaffold pairs

Homoeologous pairs were detected through identification of orthologs shared between scaffolds (Supplementary Methods 2.1). Pairs sharing a large number of orthologs, and nucleotide similarity were considered homoeologous. This was validated with other methods, including MASH distance (Ondov et al., 2016) and shared possession of duplicated BUSCO genes (Simão et al., 2015).

Phasing of subgenomes

Scaffolds were phased into A and B subgenomes using a k-mer based approach built on the approach taken by (Cerca et al., 2022). K-mers in the assembly were detected and counted using *jellyfish* (Marçais & Kingsford, 2011), with only k-mers present more than 75 times in the assembly and represented at least twice as often in one subgenome than the other carried forward. These counts were then transformed into binomial distributions. Hierarchical clustering was performed on these sets, creating a dendrogram placing scaffolds into opposing clusters, each cluster representing a subgenome (Supplementary Methods 2.1).

2.2.8 Synteny analysis

The *MCScan (Python)* (Tang et al., 2008) module of *JCVI* (Tang et al., 2015) was used to perform a synteny analysis between the two subgenomes. Custom scripts then extracted collinearity information and generated synteny plots (Supplementary Methods 2.1).

2.3 Results

2.3.1 Sequencing and profiling of read libraries

We used PacBio single-molecule real-time (SMRT) sequencing technology, Hi-C chromatin conformation capture, Nanopore long-read sequencing, and Iso-Seq RNA sequencing to generate a genome assembly of *Meloidogyne javanica* strain VW4. Following quality control and concatenation of two libraries, we obtained 2,255,922 PacBio HiFi reads totalling 35.39 Gbp (Supplementary Figure 2.1). After quality control and concatenation of Oxford Nanopore (ONT) PromethION data, we obtained 340,373 reads totalling 17.52 Gbp (Supplementary Figure 2.2). Our Hi-C library contained 375,330,537 read pairs, of which 26.20% were sufficiently unique and high quality for scaffolding with *Proximo* (Phase Genomics, WA). After demultiplexing of our Iso-Seq library, we obtained 2,506,897 full-length non-chimeric sequences which collapsed into 59,637 high quality isoforms.

Genome-wide k-mer profiling of concatenated PacBio HiFi read libraries with *smudgeplot* (Ranallo-Benavidez et al., 2020) indicated that 48% of the genome was tetraploid, 22% was triploid, and 29% was diploid (Figure 2.1a). An incomplete or hypo-tetraploid state was also indicated using a k-mer spectra approach by *GenomeScope2*, predicting a haploid genome length of 68 Mbp, and a duplication rate of 3.4 (Figure 2.1b; Ranallo-Benavidez et al., 2020). This duplication rate is similar to the value seen for CEGMA genes in other assemblies of *M. javanica* (3.68; Blanc-Mathieu et al., 2017).

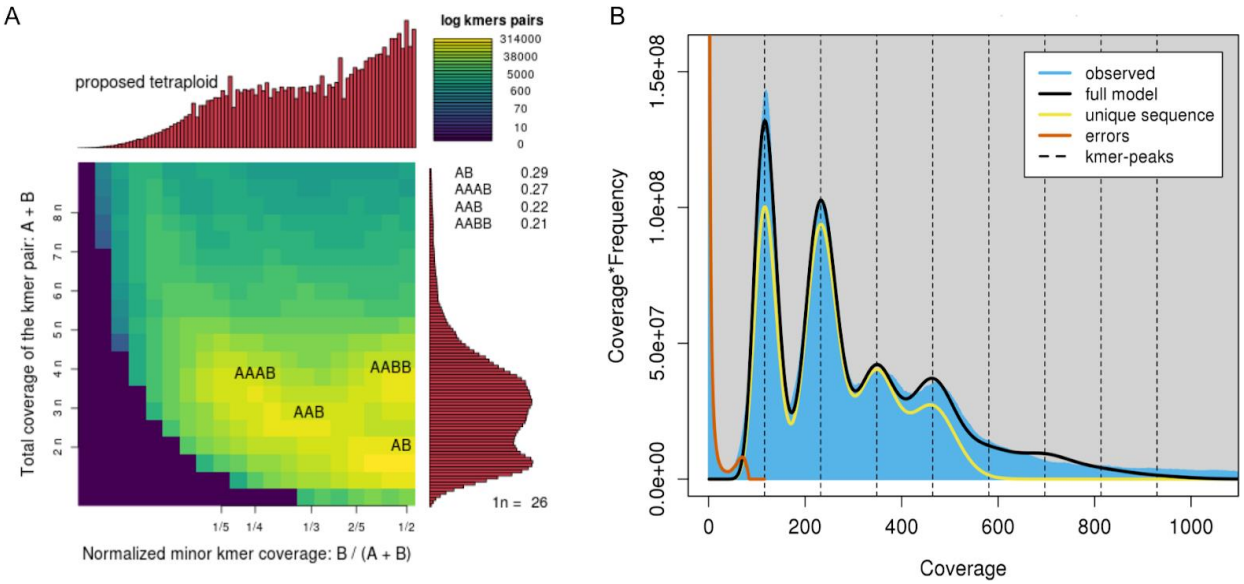


Figure 2.1: Genome profiling plots. A, *Smudgeplot* (left) proposing *M. javanica* as tetraploid, reporting the predicted percentages of ploidy levels in the genome as follows: tetraploid (48%), triploid (22%), or diploid (29%). **B**, *GenomeScope2* plot (right) showing four distinct peaks in both the predicted model of tetraploidy (black line) and in the observed k-mer spectra (blue fill). The amount of unique sequence falls to zero shortly after the fourth peak, indicating that k-mers at higher ploidies than four were mostly repetitive elements.

2.3.2 Assembly and annotation

Scaffolding and final assembly

From our draft assemblies of the PacBio data, we carried forward an iteration assembled using *HiFiasm* (Cheng et al., 2021) based on overall contiguity and comparison to expected diploid genome length. Following purging of duplicates from the PacBio assembly and scaffolding with Oxford Nanopore reads (Supplementary Methods 2.1) we obtained 37 contigs. Hi-C scaffolding with the *Proximo* pipeline identified 16 chromosome level clusters (Phase Genomics Ltd) but increased the total number of scaffolds from 37 to 66 (Supplementary Figure 2.3). Following scaffolding with Hi-C, *samba* (Zimin & Salzberg, 2022) joined some small scaffolds and fragmented the largest scaffold in the assembly, increasing the number of scaffolds from 66 to 69.

The final assembly scaffolds contain 150,545,692 bp with an N50 of 5,793,182 bp, at 30.11% GC content, and overall, 99.96% of reads in our combined PacBio HiFi library map successfully back to the assembly, indicating a high level of completeness (Supplementary

Table 2.1). Of the total assembly, 97.87% was contained in the longest 33 scaffolds (Supplementary Figure 2.4), which ranged from 898 Kbp to 9,595 Kbp in length (Figure 2.2a; Supplementary Table 2.2). These 33 scaffolds contained 99.85% of all transcribed gene models detected from our Iso-Seq data. Of the remaining 36 scaffolds, 4 were identified by *blobtools* (Laetsch & Blaxter, 2017) as likely contaminants (Arthropoda, Chordata, and Streptophyta; Supplementary Figure 2.5). Nematode scaffolds are often misclassified as Arthropoda by *blobtools*, and so have been retained in the assembly. One further contig was the *M. javanica* mitochondrial genome (22,238 bp). The remaining 31 small contigs were all less than 215 Kbp, with a mean length of 88.9 Kbp, and since they contained few identifiably functional elements (0.19% of gene models), we excluded them from the final coverage and synteny analysis.

Annotation

In total 30.46% of our assembly was identified as repetitive elements, with 4.94% identified as retroelements, 3.77% as DNA transposons, while 16.88% remain unclassified repeats (Supplementary Table 2.3). A total of 164,394 transcripts, representing 59,632 isoforms, were detected through mapping of our Iso-Seq library, of which 97% of reads mapped to the assembly. *MAKER3* detected a total of 22,433 genes, containing 227,617 exons, 10,044 5' and 5,253 3' UTRs (Supplementary Table 2.4). When running BUSCO on transcriptome settings against the eukaryote_odb10 database we find that 81.9% of genes are present (C:81.9% [S:43.5%, D: 38.4%], F:8.6%, M:9.5%, n:255). When running against the nematoda_odb10 we find 59.2% of BUSCOs (C:59.2% [S:27.5%, D:31.7%], F:2.7%, M:38.1%, n:3131).

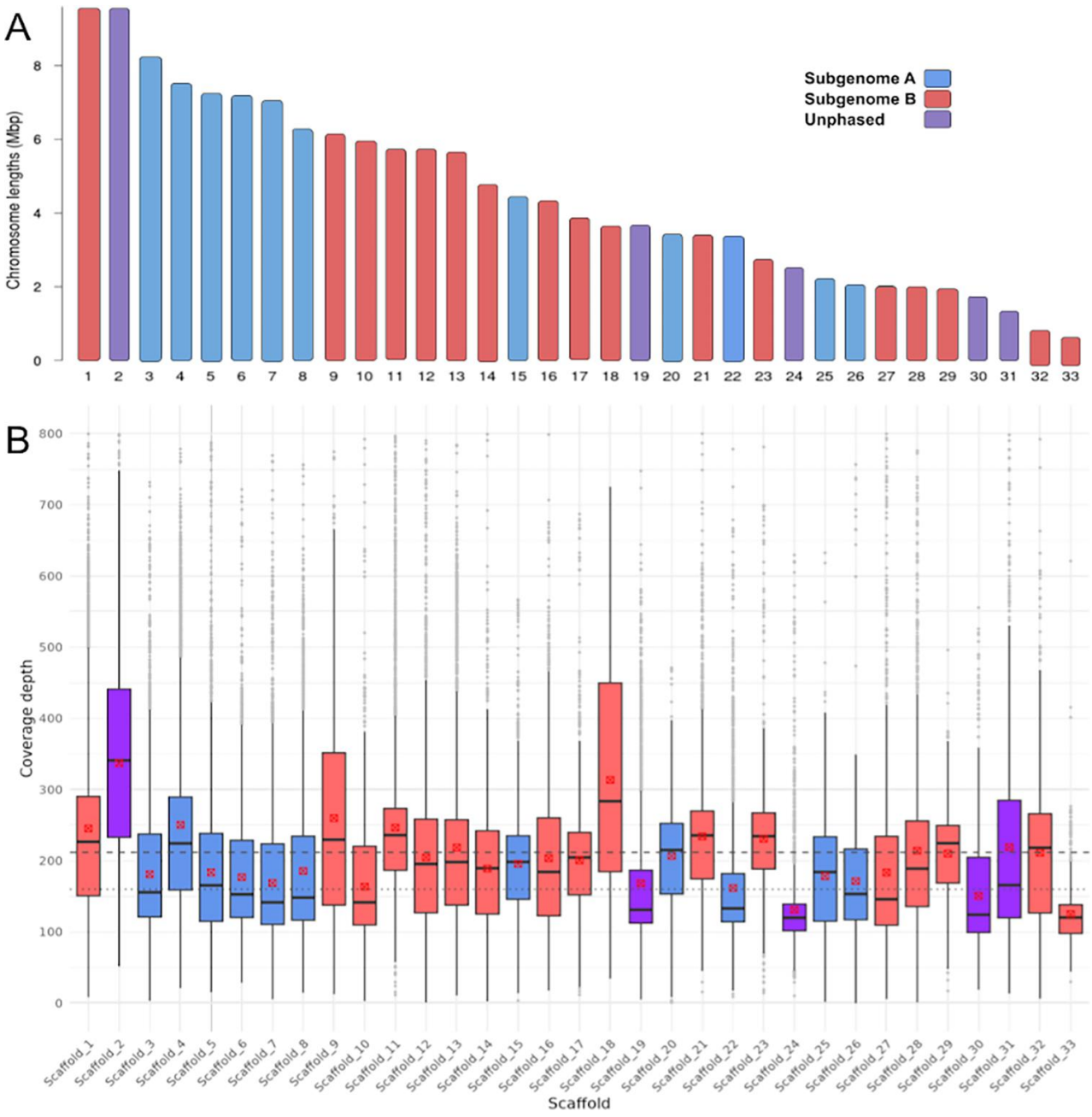


Figure 2.2: Ideogram and coverage depth of longest 33 scaffolds. Scaffolds are coloured according to phasing status; blue - subgenome A, red - subgenome B, purple - unphased. **A**, Ideogram of 33 largest scaffolds. These 33 scaffolds contain 98% the total length of the assembly, with the remaining 36 contigs being shorter than 250 Kbp, containing few gene models (0.19%), and consisting of mostly repetitive elements. **B**, Boxplot displaying the distribution of coverage depth of mapped PacBio HiFi sequences for each scaffold. Red points denote the mean of data in each box. Coverage has been limited to a maximum of 800x to exclude probable repetitive sites with anomalous coverage depth. Dashed line shows the overall mean for all coverage levels across 33 scaffolds. Dotted line shows the mode of all coverage across 33 scaffolds.

2.3.3 Core eukaryotic gene and single universal copy ortholog analysis

Analysis of the final assembly with *CEGMA (Parra et al., 2007)* detected 233 of 248 CEGMA genes (93.95%). This is a higher level of completeness than the most recent *M. javanica* assembly, and comparable with contemporary *Meloidogyne* assemblies (Supplementary Table 2.1; Szitenberg et al., 2017; Koutsovoulos et al., 2019; Susič et al., 2020; Kozlowski et al., 2021).

The average number of orthologs for each complete CEGMA gene (a proxy for ploidy) is 1.88 indicating that only ~6% of the diploid genome is unassembled or not present in the biological chromosomes. 177 complete BUSCO genes were detected when using BUSCO's eukaryote database, representing 69.5% of genes in the database (C:69.5% [S:37.3%, D:32.2%], F:13.7%, M:16.8%, n=255). Of the BUSCOs identified, 32.2% were duplicated. When using BUSCOs nematode database, we find 1556 complete BUSCO genes, representing 49.7% of genes in the database (C:49.7% [S:22.5%, D:27.2%], F:3.9%, M:46.4%, n=3131).

2.3.4 Coverage and ploidy analysis

Mean coverage depth for all scaffolds - excluding mitochondrial - was 206.6x, falling to 206.1x for the longest 33 scaffolds (Figure 2.2b; Supplementary Figure 2.6). For some scaffolds (2, 9 and 18) coverage depth was significantly more than this average, indicating collapsed regions; that is, these scaffolds represent three or four genomic copies, rather than two. Collapse is expected to arise in a polyploid assembly when homoeologous sequences are similar enough at a nucleotide level that they are inferred to be multiple homologous alleles of the same region, which are then collapsed to form the reference assembly copy (Zhang et al., 2020). Other scaffolds (19, 22, 24, 33) exhibited lower mean coverage than the assembly-wide average suggesting that they may be present as a single copy.

We then examined the coverage depth frequency distribution for each scaffold (Supplementary Figure 2.7). For a phased scaffold representing two identical homologs, the coverage depth distribution for that scaffold would be expected to have a single peak (240x). However, while predominantly single peaks were seen for some scaffolds with lower

coverage - thought to be present as a single copy (19, 22, 24, 33) - we observed two peaks for the majority of scaffolds (Figure 3a & 3c; Supplementary Figure 2.7). Two peaks would be expected if homologs are not identical and similarity is disrupted by indels, with the second peak representing the reduced coverage of the hemizygous region (~120x). Following this logic, number of peaks in the plot would indicate a respective number of biological alleles mapping to one copy in the genome assembly. For scaffold 2, which is over-represented in mean coverage depth, four clear peaks are present (Figure 2.3) indicating that more than two alleles map to the scaffold. This likely resulted from exclusion of the scaffold's homoeolog from the assembly, leading to an assembly collapse as noted above. Nevertheless, the presence of 4 peaks is consistent with the presence of polymorphisms between homologs as well as between homoeologous pairs.

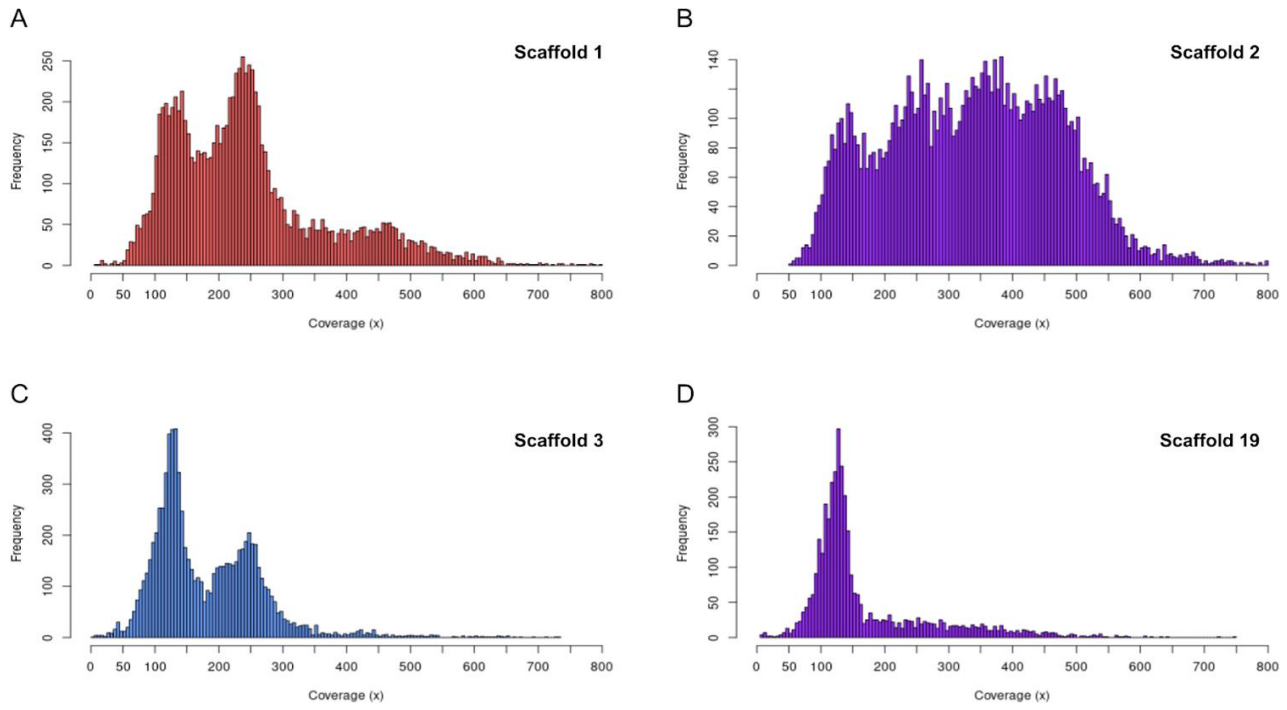


Figure 2.3: Coverage depth frequency distributions of scaffolds 1, 2, 3, and 19. The X-axis represents coverage depth of mapped PacBio HiFi reads and the y-axis represents frequencies of those coverages by individual bases. Colour indicates phase status: red for subgenome B, blue for subgenome A, and purple for unphased scaffolds. Bin size = 5. **A**, Scaffold 1, top left, displays two main peaks of coverage with a small tail suggesting short collapsed regions. **B**, Scaffold 2, top right, displays four peaks of coverage, indicating that much of this scaffold is collapsed and four copies are mapping to it. **C**, Scaffold 3, bottom left, shows two peaks and very little tail, indicating two copies mapping and little to no

assembly collapse. **D**, Scaffold 19, bottom right, shows only one peak at ~120x coverage, suggesting that only one copy maps to this scaffold.

To examine the ploidy distribution in another way, we plotted a sliding window of coverage across each scaffold (Supplementary Figure 2.8). In support of the coverage depth distributions, the most frequent outcome was that scaffolds in the assembly show two layers of stratification in coverage at a constant proportional depth, indicating two copies distinguished by indels between them. This pattern was strongest for scaffolds with two coverage peaks, particularly phased homoeologs (discussed below, Figure 2.3a and 2.3b; Supplementary Figure 2.7). Scaffold 2 and scaffold 18, which are over-represented in sequence depth, contain regions of four levels of coverage depth covering much of the length of both (Figure 2.3b; Supplementary Figure 2.8a and 2.8d). This increased amount of stratification at proportionally higher coverage depths (360x and 480x) reveals assembly collapse, where three or four copies, respectively, map to a single site. Together these coverage depth results suggest that the *M. javanica* assembly represents two homoeologous subgenomes (~85% of total length), with only 13% of the assembly unphased or collapsed.

2.3.5 Identification of homoeologous pairs and phasing of subgenomes

Given the diploid nature of the assembly, which represents each subgenome as a single copy, we expected to find scaffolds from these subgenomes present in homoeologous pairs. Through detection of shared orthologs (Supplementary Methods 2.1) twenty pairings were identified between the longest 33 scaffolds of the assembly, each sharing between 20 and 351 CDS orthologs (Figure 2.4; Supplementary Table 2.5). Alternative methods of identifying homoeologous pairs were corroborative (Supplementary Table 2.6). Some pairs were not mutually exclusive, and exhibited CDS links to scaffolds outside of the primary pair, suggesting translocation and syntenic changes between them. Four scaffolds were excluded from pairings including scaffold 2 which showed high amounts of collapse and scaffolds 19, 24, 30, and 31 which are relatively small and/or present as single copies (Supplementary Figure 2.7 & 2.9; Supplementary Table 2.2).

In order to assign contigs to a subgenome, we used a modified version of an ancestral k-mer spectra analysis (Supplementary Methods 2.6; Cerca et al., 2022). This approach is based on the premise that the allopolyploid's subgenomes possess repeats that diverged in the two parental genomes before hybridisation. This results in a distinguishable signature in each subgenome's k-mer spectrum and allows us to determine the parental species from which a given sequence descends. We successfully phased 85.39% of the assembly into A or B subgenomes (Figure 2.2a; Supplementary Figure 2.9; Supplementary Table 2.2). Scaffold 2, which exhibits extensive assembly collapse, did not phase using these methods. Some smaller scaffolds that did not phase by k-mer based methods were later assigned to a subgenome based on the phase status of their opposing homoeologous scaffold (Supplementary Table 2.2). Of the total length of the final assembly, 39.15% was assigned to subgenome A and 46.24% was assigned to subgenome B, leaving only 14.61% unassigned. Nucleotide similarity between the subgenomes was estimated at 86% for whole subgenomes and 91% between only CDS regions. Allelic divergence between alleles mapping to either subgenome was estimated, with subgenome A estimated at 97.1% and subgenome B at 97.7% (Supplementary Methods 2.7).

2.3.6 Subgenomic synteny analysis

Comparison of annotation of scaffold pairs assigned to subgenomes A and B revealed both long regions of synteny and large structural differences between subgenomes (Figure 2.4). Scaffold 13 and scaffold 15 are syntenic along almost the entire length of the shorter homoeolog and share high nucleotide similarity throughout. Similarly, scaffold pairs 20 and 23, as well as 17 and 25, share long syntenic blocks of shared CDS with high nucleotide similarity.

Scaffolds 7 and 10 are syntenic for almost half of their length whilst the remaining sequence lengths share no synteny and have very low nucleotide similarity. Synteny analysis and scaffold comparison suggest that chromosomal fragmentation has occurred. For example, scaffold 8 of subgenome A shows long collinear blocks with scaffolds 16, 28, 29, and 33 of subgenome B. Re-examination of Hi-C and ONT long-read scaffolding as well as manual inspection of read mapping support this fragmentation. Scaffold 1 of subgenome B shares syntenic blocks with scaffolds 5, 6, and 8 of subgenome A, in an order that suggests chromosomal structural differences between the subgenomes. Many phased

scaffolds also exhibit small amounts of extra-pair synteny indicating numerous small translocations throughout the genome.

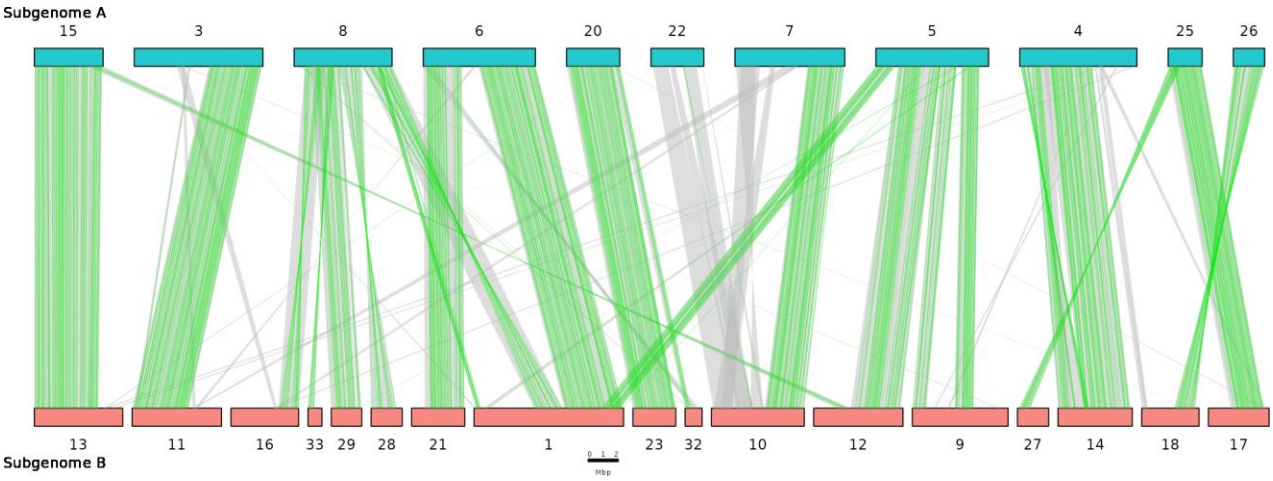


Figure 2.4: Macrosynteny analysis between subgenomes. Glyphs along top and bottom represent scaffolds assigned to either subgenome A (blue) or subgenome B (red). Green lines mark locations of synteny between transcribed genes identified from mapping of Iso-Seq sequences. Grey lines mark locations of synteny between genes identified through *MAKER3* gene prediction. Synteny and collinearity were identified using the *MCScan* module of *JCVI* using Iso-Seq informed transcriptional annotation. Scaffolds that could not be assigned to a subgenome are not shown.

2.4 Discussion

2.4.1 Assembly of the allopolyploid genome of *M. javanica*

We have used long-read sequencing and modern bioinformatic approaches to assemble and phase the allopolyploid genome of the plant pathogenic nematode *Meloidogyne javanica*. Our goal was to assemble a contiguous diploid assembly for *M. javanica*, representing the A and B subgenomes separately. Our current diploid assembly (150,545,692 bp) is, as expected for a tetraploid, approximately half the 297 ±27Mb measured by flow cytometry (Blanc-Mathieu et al., 2017; Szitenberg et al., 2017). This total length is representative of both A and B subgenomes except for collapsed regions where sequence for both subgenomes A and B is considered together. Because some regions of the assembly have been shown by k-mer profiling and coverage analysis to be present in

less than four copies (Figure 2.1; Supplementary Figure 2.8), splitting the subgenomes into their component copies (A1, A2, B1, B2) to create a tetraploid assembly would not be expected to completely double the length.

Annotation

We annotated our assembly using both *ab initio* feature prediction algorithms and mapping of the full-length transcript sequences (Supplementary Table 2.4). The total number of genes predicted - 22,433 - is within the range expected for a MIG species and is comparable to previous *M. javanica* assemblies (Blanc-Mathieu et al., 2017; Szitenberg et al., 2017).

BUSCO scores for *Meloidogyne* are consistently lower than those of more widely studied organisms, and the number of genes we have detected (Supplementary Table 2.4) in the assembly is consistent with what has been found for other *Meloidogyne* species (Supplementary Table 2.1; Blanc-Mathieu et al., 2017; Szitenberg et al., 2017; Koutsovoulos et al., 2020). The paucity of established protein databases for less frequently investigated genera limits the accuracy of prediction-based annotation methods. With increased availability of sequence resources for plant parasitic nematodes, it should soon be possible to develop a more appropriate set of core genes.

Evidence for a chromosome-scale assembly

Previous genomic assemblies of *M. javanica* have contig counts numbering in the thousands (Blanc-Mathieu et al., 2017; Szitenberg et al., 2017). In our current assembly, more than 99% of reads in our PacBio HiFi libraries map to 33 large scaffolds (Figure 2.2b). We propose that most of the 33 scaffolds represent full-length or nearly full-length chromosomes. Cytological examination in *M. javanica* indicates that the chromosome number ranges from 42-48 due to variation between isolates (Triantaphyllou, 1985; Eisenback & Triantaphyllou, 1991). Thus, we would expect 21-24 scaffolds in our assembly. The discrepancy between scaffold number and cytological observations could be due to an imperfect assembly or failure to identify very small chromosomes in the cytological studies. We have employed several independent scaffolding softwares, Hi-C chromatin contact mapping, and the manual examination of long-read mapping to contig termini, and see no evidence to support fusing additional contigs. Additional molecular and cytological studies may be required to resolve these differences.

Many chromosome-scale assemblies identify telomeres as defining the range of their scaffolds, yet we did not identify canonical telomeric repeats at scaffold termini. Additionally, we were also not able to identify a homolog of *C. elegans* telomerase (*trt-1;* ACC: NM_001373211.4) in our assembly. This may suggest that non-standard telomere processes might be operating in root-knot nematodes as has been found for some other animals (Mason et al., 2011; Pardue & DeBaryshe, 2011).

2.4.2 Genetic variation in *M. javanica* is dominated by indels

We present the assembly as a diploid representation. Read depth analysis of individual scaffolds indicates that homologs are not identical and indels are frequent. Additionally, some scaffolds appear to be present as a single copy suggesting that one of the homologs may have been lost. We identified many indels that delete or disrupt one or more copies of coding sequences, indicating that *M. javanica* is no longer genome-wide tetraploid either in copy number or functionality. A higher propensity for indel accumulation has been frequently seen in hybrid parthenogenetic species (Jaron et al., 2021) and partial return to lower ploidy is a characteristic of many polyploids.

We find that 85.39% of our diploid assembly consists of one copy of each subgenome in homoeologous pairs, with two copies mapping to phased scaffolds and four copies mapping to collapsed regions. We successfully phased much of our assembly into subgenomes using k-mer signatures, enabling for the first time initial genome-wide comparison of MIG A and B subgenomes. Some scaffolds, notably scaffold 2 and some of the short scaffolds, could not be assigned to subgenomes. Scaffold 2 displays four levels of stratification in its coverage (Supplementary Figure 2.8) and four peaks in its depth distribution (Figure 2.3b) indicating that the four copies (A1+A2 and B1+B2) are almost entirely collapsed into a single scaffold. We suggest that the homoeologous chromosomes represented by scaffold 2 may be too similar in sequence to assign to a subgenome by k-mer analysis due to a possible ancient homogenization event. For the smaller scaffolds, the failure may have been due to their small size because they did not contain enough relevant k-mers. Some of the small scaffolds were later assigned to a subgenome using transcript alignments. No misassembly or erroneous scaffolding of the unphased sequences was detected through either programmatic or manual methods.

The allotetraploid genome of *M. javanica* demonstrates extensive synteny between the A and B subgenomes with most genic regions represented in four copies, as detected from coverage stratification and ploidy profiling. We would expect however that there would be differences between the subgenomes, including indels and other structural variation, as this is typically observed between different species and the MIG have a hybrid origin. In accordance with this, we observed substantial structural variation between subgenomes A and B, including insertions, deletions, and translocations (Figure 2.4).

2.4.3 Loss of synteny and fragmentation

We observe regions of the *M. javanica* genome where synteny between paired chromosomes is disrupted. One reason for this could be that the parental species (A and B) had genomes in which the non-syntenic regions had diverged by translocation, insertion, or deletion. Upon hybridisation to create the allopolyploid MIG these diverged regions form the end of synteny blocks. An alternative explanation is that these changes happened after the hybridisation event in the tumultuous process of genome stabilisation immediately following it (Edger et al., 2018; Emery et al., 2018). Hybridisation, polyploidisation, and the loss of meiosis are processes often associated with rapid genomic change (Balloux et al., 2003; Edger et al., 2018; Carlton et al., 2022) and the unique MIG species allopolyploid genomes we are currently studying may represent different balances between these forces. We observe 11 chromosome-scale scaffolds in subgenome A and 17 in subgenome B despite the clear synteny throughout these two subgenomes. Several scaffolds in subgenome A contain blocks of genes with regions that are syntenic to different subgenome B scaffolds. Similarly, there are cases where syntenic blocks in subgenome B are present on different scaffolds in subgenome A. Together these differences suggest that ancestral chromosomal fission, or fusion events or other types of exchange have occurred. *Meloidogyne* species, like other nematodes, have holocentric chromosomes (Triantaphyllou, 1981). Genomes with dispersed centromere structure are predicted to better tolerate chromosome fragmentation and fusion (Carlton et al., 2022), as are species with ameiotic mechanisms of reproduction. The observed differences in copy number of chromosomes between isolates of *M. javanica* may be additional evidence for tolerance of chromosome fragmentation/fusion.

The majority of published assemblies of allopolyploids come from vascular plants, where polyploidy might have shaped the genomes of around 70% of species (Masterson, 1994). Many allopolyploid plant genomes, however, show a higher level of synteny and

structural conservation than we observe for *M. javanica* (Cerca et al., 2022; Shen et al., 2022; Zheng et al., 2022). Similarly the few available chromosomal allopolyploid animal genome sequences available (Du et al., 2020; Kuhl et al., 2022) do not show extensive deletions and chromosomal fissions as does our genome assembly. Unlike other allopolyploid species with chromosomal genome sequences, *Meloidogyne javanica* reproduces by obligatory mitotic parthenogenesis (apomixis) and this lack of meiotic chromosome pairing may allow greater structural divergence and tolerance for the decay of synteny. We note that genomes from species not able to reproduce by meiosis are currently rare (Jaron et al., 2021) and suggest that much more substantial genomic work on a range of species with different reproductive modes and ploidy levels will be required to reveal the diverse mechanisms shaping these genomes. It is apparent however that the changes surrounding allopolyploidy have contributed to the gene content, heterozygosity, and copy number throughout the *M. javanica* genome and these processes of locally fixed heterozygosity or rediploidisation may contribute extensively to adaptive functional variation (Dodsworth et al., 2016).

2.4.4 Regional homogenization of subgenomes

Some regions of the genome assembly do not phase into A and B homoeologs due to very low divergence between the four gene copies. This could be explained either by the loss of whole chromosomes, compensated by the duplication of the remaining chromosome, or mitotic recombination (gene conversion) between homoeologs (Harris et al., 1993; Mansai & Innan, 2010). It is unclear from this single genome how asexual recombination contributes to shaping the diversity of *M. javanica*, however this has been suggested in previous MIG genomic studies (Blanc-Mathieu et al., 2017; Szitenberg et al., 2017) and may be further elucidated by our ongoing molecular evolution and population genomic studies.

It is possible that the initial tetraploidisation of *M. javanica* will buffer against deleterious phenotypic consequences of indels. It has been argued for angiosperms that this rediploidization process can contribute to adaptive species divergence by providing genomic and transcriptomic diversity (Hollister, 2015; Dodsworth et al., 2016). Other mechanisms of adaptive divergence may also operate at the same time. The increase in gene copy number created by polyploidization gives the potential for functional gene divergence by neo- or sub-functionalization, as well as adaptive phenotypes driven by copy number loss (Paquin & Adams, 1983; Otto & Whitton, 2000; Zörgö et al., 2013). Adaptation by gene copy number

variation has already been reported in *M. incognita* (Castagnone-Sereno et al., 2019) and it may be that genomic copy number variation more broadly is a major source of functional genetic variation in the MIG.

2.4.5 A genomic framework for RKN functional and diversity studies

In this paper we present a highly contiguous, annotated, and phased genome assembly of the allotetraploid plant pathogenic nematode *M. javanica*. This genome assembly will provide many tools for diverse investigations by plant pathologists and nematologists in addition to aiding our understanding of the origins and diversity of *M. javanica*. It will also serve as a reference for investigating genome structure and pathogenicity in other *Meloidogyne* species. The contiguous nature of this genome and the high-quality annotation will facilitate RKN functional studies since transcripts can be mapped accurately to the annotated subgenome. This allows consideration of copy number variation, which may be an important component of functional variation in these species. Progress is being made by many groups in understanding the basis of nematode virulence and the key loci involved (Pogorelko et al., 2019; Kihika et al., 2020; Song et al., 2021). In such cases even light coverage sequencing of field isolates mapped to a high-quality genome assembly could give valuable information to commercial growers about the likely pathogenicity of those strains (Sellers et al., 2021).

Chapter 3: A comparative analysis of subgenomes in allopolyploid root-knot nematodes of the *Meloidogyne* genus

Michael R Winter[1]

[1] School of Natural Sciences, University of Hull, Kingston-upon-Hull, UK.

Correspondence: mrmrwinter@gmail.com

3.0 Abstract

The interaction of homeologous subgenomes in asexual animal species is understudied. The *Meloidogyne incognita* group contains several species exhibiting this asexual allopolyploid system, which combined with novel genomic resources, can enable investigation of this uncommon genomic architecture. We performed an exploratory analysis of the genomes of the *Meloidogyne incognita* group and *Meloidogyne luci* utilising a recently produced phased genome assembly of the root-knot nematode *Meloidogyne javanica,* in order to interpret and compare subgenome-specific molecular evolutionary histories. In addition, we perform synteny analysis for each subgenome between *M. javanica* and other species, and comparison of codon frequency, homogeneity of bases, and k-mer spectra based ploidy analysis. Our objectives were to identify dominant subgenomes, assess phylogenetic relationships of subgenomes, evaluate synteny conservation, and attempt to detect evidence of DNA methylation. Using a phased assembly as a reference, we successfully phased contigs of other species, and lifted over gene annotations, before performing differential selection tests. Our results indicate that relative rates of non-synonymous mutation differ significantly between subgenomes among most species. Subgenome A shows a slight dominance over subgenome B, although no definitive evidence of subgenome scale dominance was observed. Selection pressures acted oppositely on homoeologous genes at some loci of *M. javanica*, potentially contributing to subgenome dominance and fractionation. A considerable amount of synteny and collinearity is conserved across species. Phylogenetic analysis reveals that subgenomes are more closely related to orthologous subgenomes in other species than to their homoeologous counterparts within the same species, confirming homologous origin of subgenomes of these species. Notably, *Meloidogyne luci* follows this pattern, suggesting a shared origin and earlier divergence. Additionally, evidence of DNA methylation was detected in one species for the first time. Our study provides new insights into subgenome dynamics and evolutionary relationships within these species, highlighting the role of selection and genomic conservation in shaping allopolyploid genomes.

3.1 Introduction

Allopolyploidy originates from the hybridisation of two species and subsequent retention of path parents' genomes in the offspring lineage (Kostoff, 1939; Ma & Gustafson, 2005; Schoenfelder & Fox, 2015). For example, two diploid (AA and BB; 2n = 10) species hybridise to form an allotetraploid species (AABB; 2n = 4x = 20). Many allopolyploid species have been identified, including the common carp (Kuhl et al., 2022; Chen et al., 2023) and the African clawed frog (Knytl et al., 2023), as well as several important agricultural and commercial species, such as Arabica coffee (Salojärvi et al., 2024), cultivated wheat and cotton (Feldman & Levy, 2012; Chen et al., 2020), and rapeseed (Cao et al., 2023).

Allopolyploidy is clearly a powerful mechanism through which genomic diversity can be generated, but the details of how a genome changes and the evolutionary forces acting on it following hybridisation are unclear. Allopolyploid systems allow the study of selection and evolution in a unique genomic system, from which insights have been demonstrated to have deployable benefit in agriculture, aquaculture, and medicine (Sattler et al., 2016; Zhou & Gui, 2017; Fox et al., 2020; Anatskaya & Vinogradov, 2022). Many species important to civilisation are allopolyploid, including many species of agriculturally essential plants, with many more predicted to have undergone hybridisation and polyploidy events in the past (Zhang et al., 2019; Heslop-Harrison et al., 2022).

The complex nature of *Meloidogyne incognita* group (MIG) genomes provides us with a case study to investigate the evolution of allopolyploid lineages and their genomes, observe how homeologous subgenomes evolve within an asexual species and how homologous subgenomes evolve across a group.

3.1.1 Genomics and evolution of subgenomes in allopolyploids

Allopolyploidy

Allopolyploids can form through multiple pathways, one of which is the triploid bridge. Under this method, hybrid triploids arise by chance through interspecific crosses which then self fertilise or backcross to produce stable allotetraploids (Yamauchi et al., 2004; Li et al., 2022). Another is the one-step method, where allopolyploids are produced through the interspecific crossing of two unreduced gametes (Oleszczuk & Lukaszewski, 2014; Mason & Chris Pires, 2015; Clo et al., 2022). Hybridisation of individuals with already non-standard genomes can also lead to allopolyploidy, such as the crossing of autopolyploid species

(Duan et al., 2024), or a significant incongruence between karyotypes of the parent species leading to a stabilisation of a new even-ploidy karyotype (Ramsey & Schemske, 1998; Martin et al., 2020). Allopolyploidy is believed to be uncommon in nature due to factors preventing stability and success of the hybrid and its genome, such as pre- and post-zygotic barriers like incompatibility in metaphase from differing karyotypes, and increased copy number of many genes leading to lethal overexpression, as well as ecological factors such as reduced fitness in comparison to the sympatric parent species. However, if the hybrid can overcome these hurdles, allopolyploidy confers many opportunities for adaptation. Allopolyploidy typically increases the complexity of a genome and gives an immediate increase in genome redundancy, and through it diversity and heterozygosity.

Redundancy occurs where there are duplicates of a functional gene. One copy is conserved, and the others accumulate mutations (Gibson & Wagner, 2000; Krakauer & Plotkin, 2002). This occurs in polyploids due to the immediate increase in copy number following genome duplication and is considered a mechanism by which polyploids succeed when sympatric to the ancestral population (Comai, 2005). Redundancy is gradually reduced through a process of ongoing gene loss and neofunctionalisation called gene fractionation, which reduces genome redundancy and alleviates gene dosage imbalances, and is explained in greater detail later in this section. Genome redundancy reduces the probability of fixed deleterious recessive alleles and allows finer regulation of dosage-based gene expression (Osborn et al., 2003). It also increases the possibility of sub- and neofunctionalization (Comai, 2005; Birchler & Yang, 2022), where duplicate copies can both contribute to gene expression, or the duplicate copies can gain novel function. The potential for increased dosage could result in a phenomenon called hybrid vigour, wherein interspecific hybrids display increased size, fecundity, and development speed than their progenitors (Birchler et al., 2010; Chen, 2010).

Asexuality

Asexuality is a state where an organism reproduces parthenogenetically, cloning the gamete content either pre- or post- meiotically (Suomalainen, 1950; Bell, 2021; Jaron et al., 2021). Asexuality and parthenogenesis occur frequently in allopolyploids, which is suggested to be a result of genome redundancy increasing tolerance to accumulation of recessive alleles (Comai, 2005; Neiman et al., 2014). In some species, asexuality has been shown to be a result of hybridisation and be correlated with phylogenetic distance of the progenitors (Choleva et al., 2012; Janko et al., 2018). Coupled with reproductive incompatibility with the parent species enforcing reproductive isolation and preventing

outcrossing and the beneficial effect of hybrid vigour, this leads to the rapid establishment of sympatric allopolyploid species following successful and stable hybridisation.

Asexuality is expected to initiate Muller's ratchet, a situation wherein absence of meiosis and recombination leads to an accumulation of deleterious mutations (Muller, 1932), however it has been shown that this is not always the case in allopolyploids, likely due to homoeologous recombination (Pellino et al., 2013; Maciver, 2016).

In fact, parthenogenetic polyploids are considered to be more tolerant to some of the negative mutational effects of asexuality such as gene deterioration and deletion thanks to increased genome redundancy (Comai, 2005). Amphimixis occurs in the wider *Meloidogyne* genus (Figure 3.2), notably in *Meloidogyne hapla*, and males of some apomictic MIG species have been identified (Triantaphyllou, 1963). If in rare cases the clonal MIG species were able to cross with rarely produced males, the ensuing recombination could offset some of the mutational load and increase the stability of the lineage (Liu et al., 2007; Maccari et al., 2013).

Asexuality also permits the coexistence of homologous polyploid cytotypes within an ecosystem (Kao, 2007). This is pertinent to the MIG, where several species have diversified and stabilised with homologous subgenomes. Asexuality also means that subgenomic structure is conserved through generations.

The fate of allopolyploid genomes

Subgenomes in an allopolyploid often exhibit a phenomenon called subgenome dominance. Subgenome dominance refers to the trend of one subgenome being primarily conserved and retaining much of its ancestral function, whilst the other mutates or degrades. Higher dosage is observed from the dominant subgenome, as well as a higher number of homeologous conversion events from dominant to non-dominant (Alger & Edger, 2020). The terms used to ascribe roles to either subgenome are not to be confused with the dominant/recessive terminology used in Mendelian genetics. Throughout this work, when discussing subgenome dominance we use the terms dominant and non-dominant to refer to this trend. Subgenome dominance can occur at the whole subgenome level rapidly following hybridisation, but this is dependent on hard thresholds where a mismatch between genetic machinery can be highly deleterious or even lethal. Examples of this are a mismatch between codon usage and tRNA specificity, as well as mismatches between regulatory and metabolic networks (Osborn et al., 2003; Birchler & Veitia, 2010; Alger & Edger, 2020).

Although observable at a subgenome level as a trend, subgenome dominance acts primarily at a gene level. Individual genes on one subgenome are prioritised and conserved, whilst orthologous copies mutate and accumulate mutations (Bird et al., 2018; Edger et al., 2018; Alger & Edger, 2020). This causes a higher rate of gene fractionation on the non-dominant subgenome. In species undergoing recombination - either automicts or apomicts undergoing homoeologous conversion - dominant regions can become linked, increasing one subgenome's dominance. As the non-dominant subgenome degrades, chromosomes can break and be lost, eventually leading to diploidisation (Cheng et al., 2018; Schiavinato et al., 2021; Bird, 2022).

Diploidisation is the process of return to two genomic copies (2n = 2x) and restoration of bivalent pairing and disomic inheritance following a polyploidy event (Ma & Gustafson, 2005). Gradual diploidisation can be adaptive, enabling polyploids to manage detrimental increased gene dosages post whole genome duplication (Li et al., 2021), and enabling a return to sexual reproduction and recombination providing sex linked loci have not deteriorated (Lu et al., 2021; Toups et al., 2022).

There are two primary and non-exclusive mechanisms through which diploidisation is possible: gene fractionation and cytological diploidisation (Ma & Gustafson, 2005). Fractionation and sub-, neo- and nonfunctionalisation of genes are thought to be the leading factors promoting diploidisation. Over time fractionation can lead to functional diploidy, where the genome is physically diploid, but all functional genes are located on a diploid pair of copies, increasing tolerance of chromosome changes or loss. This results from subgenome dominance interactions. Following polyploidy, the increased copy number and gene dosage must be managed, which generally results in one copy being conserved and remaining functional: dominant. Copies of genes on the non-dominant subgenome can have one of three outcomes: (1) nonfunctionalisation of all but one copy, (2) semi-degeneration or subfunctionalisation of both copies resulting in the dosage of a single functional copy, or (3) neofunctionalisation, wherein the redundant copies evolve an altered function (Lynch & Conery, 2000). Alongside these three outcomes it is suggested that in some cases where increased dosage is adaptive, particularly in asexual polyploids, multiple copies can be maintained (Wendel, 2000). The presumed fate of the majority of duplicated genes is nonfunctionalisation and loss (Lynch & Conery, 2000), leading to degradation of the non-dominant subgenome. Cytological diploidisation is another mechanism exclusive to automicts where chromosomal and genetically balanced gametes are produced in a

polyploid through meiosis. These progeny have half the chromosome complement of the parent (Moore, 2002; Ma & Gustafson, 2005; Cifuentes et al., 2010).

The fate of subgenomes in an asexual allopolyploid can differ across species, and diploidisation is not always a result. Both subgenomes can be conserved in the hybrid and graduate towards autopolyploidy over time through recombination in automicts and gene conversion in apomicts (Comai, 2005; Bomblies et al., 2016; VanBuren et al., 2020; David, 2022; Mortier et al., 2024).

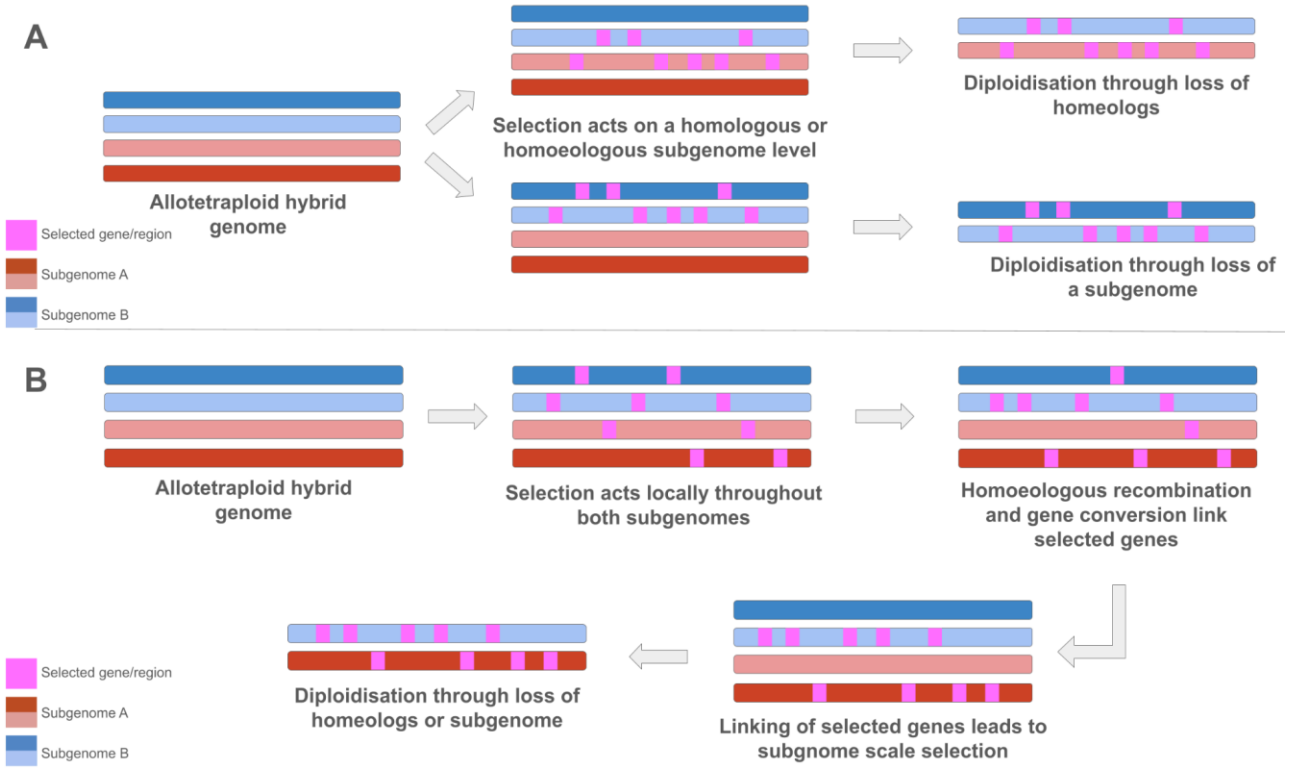


Figure 3.1: Two possible routes of diploidisation through gene fractionation. A, selection favours one subgenome (whole subgenome dominance) or one homolog of each subgenome (homoeolog dominance) over the other. Fractionation and gene loss occurs on the non-conserved copies eventually resulting in chromosome loss and a return to diploidy. This can result in haplotypes similar to one parent, or a paleopolyploid diploid haplotype representing both parents. **B,** selection acts on a more local level (local subgenomic dominance). In this situation, homoeologous recombination or gene conversion would be required to consolidate and link conserved regions onto single copies, of which the deteriorating opposing copy would subsequently be lost. Alternatively, local dominance could delay diploidisation and facilitate more opportunities for neo- and subfunctionalisation, offsetting the shift toward diploidy and aiding in establishing stable allo-or autopolyploidy.

Differences in selective pressures across and between subgenomes likely lead to different outcomes. For example, selection and dominance may act primarily on one subgenome post hybridisation, leading to an accumulation of mutations and subsequent fractionation and degradation of the non-dominant subgenome. Alternatively, selection and dominance can occur on a local level between subgenomes. This could begin on a gene by gene basis, then form larger regions as homeologous recombination generates linked blocks. In cases of local dominance, both subgenomes would be preserved for longer, and instead of diploidisation the result can be stabilisation of polyploidy or aneuploidy (Mandáková et al., 2016; Mortier et al., 2024), and an increased adaptive potential through maintenance of genomic redundancy in the non-selected genes (Comai, 2005).

3.1.2 Root-knot nematodes and the *Meloidogyne incognita* group

Root-knot nematodes (RKN), *Meloidogyne*, are a genus of important agricultural plant pathogens that infect nearly all species of angiosperm, including all commercial crop plants (Perry et al., 2009; Jones, A. Haegeman, et al., 2013). They are found on all continents except Antarctica, and are responsible for an estimated loss of 12% of the total yield, valued at $157 billion annually (Singh et al., 2015; Bernard et al., 2017; Talavera-Rubia et al., 2022), as well as contributing to agricultural losses in famine-prone regions (Barker et al., 1985; Feyisa, 2021).

Meloidogyne contains over ninety species, divided into several clades (Adams et al., 2009; García & Sánchez-Puerta, 2015; Szitenberg et al., 2017; Álvarez-Ortega et al., 2019). Some of the most damaging and impactful species are very closely related and belong to the same group within Clade I, the *Meloidogyne incognita* group (MIG; Trudgill & Blok, 2001; Bebber et al., 2014). Named for one of its member species, *M. incognita*, and also containing *M. arenaria* and *M. javanica*, the MIG represent the most damaging species of RKN. Members of the MIG are found in all tropical climates around the world and it is speculated that their range may be allowed to expand alongside rising global temperatures (Reilly et al., 2001; Bebber et al., 2013; Elling, 2013; Campbell et al., 2023). *M. luci* is another highly impactful species that is found around the world (Carneiro et al., 2014; Gerič Stare et al., 2017). It is not included in the MIG classification but is very closely related to the group based on mitochondrial and ribosomal phylogenies (Álvarez-Ortega et al., 2019). Recently,

a genome assembly of *M. luci* was published, but has not yet been widely scrutinised or utilised in molecular or evolutionary studies (Susič et al., 2020).

To prevent or treat infection, several approaches are possible including treating the field with nematicides, introducing nematicidal bacterial or fungal cultures, soil solarisation, or trap cropping, which are effective to varying degrees (Ralmi, Khandaker, et al., 2016) but it is near impossible to completely remove a MIG RKN infection once established. An alternative solution is the breeding and development of RKN resistant plant cultivars, of which several have been developed (Milligan et al., 1998; Narayanasamy, 2002; Shilpa et al., 2022), but progress is stymied by a lack of genomic resources and knowledge of the MIG. As MIG lineages are discovered that have evolved the ability to evade these engineered resistances, it is more pressing than ever that genomic resources are generated and studied, and mechanisms of adaptation and genome evolution are understood.

Genomic investigation into many clade I species (Figure 3.2), including the MIG, has been made difficult by their highly non-model genomes. There is strong evidence to suggest that many clade I species, and all of the MIG, are descendants of a hybridisation event between ancestor species (Blanc-Mathieu et al., 2017; Szitenberg et al., 2017). As such many extant species exhibit polyploid genomes, with some species, including *M. arenaria*, *M. incognita*, and *M. javanica*, reproducing exclusively through apomixis; mitotic parthenogenesis (Trudgill & Blok, 2001).

With the falling expense of long-read sequencing technology many studies are now able to generate genome assemblies and their associated annotations with higher contiguity and accuracy (Koren & Phillippy, 2015; Rhoads & Au, 2015; Jain et al., 2018; Wang et al., 2019; Logsdon et al., 2020), leading to an opportunity to investigate the genomics of the MIG with a greater resolution than ever before (Koutsovoulos et al., 2020; Susič et al., 2020; Winter et al., 2024).

3.1.3 Aim and objectives of the study

The aim of this study is to utilise the complex genomic system found in the MIG to infer the evolutionary processes involved in hybridisation and polyploid formation, as well as the behaviour of subgenome dominance and diploidisation in non-meiotic asexual species. To facilitate this, our overall objective is to perform a comparative and phylogenetic analysis of subgenomes from the species *Meloidogyne javanica, Meloidogyne arenaria, Meloidogyne*

incognita, and *Meloidogyne luci*. We previously assembled the genome of *M. javanica* and phased it into subgenomes (Chapter 2; Winter et al., 2024). Now, we perform comparative molecular analysis between the subgenomes in an effort to determine if sites or subgenomes have been under differential selection, and see if there are clear and definable differences in evolutionary patterns between subgenomes. In addition to this, we use whole genome alignment approaches to phase two recent *M. incognita* group (MIG) species assemblies, and their close relative *M. luci,* lift over annotations to identify genic regions, and place their respective subgenomes in a phylogeny for the first time. Following this we compare synteny, relative mutation rate, ploidy, and codon usage across the species, as well as attempting to detect methylation, which is thought to be a route through which allopolyploids manage genomic shock.

3.1.4 Interpreting molecular and genomic differences

With the availability of genome assemblies from several closely related allopolyploid species we can investigate several things. First, can we detect subgenome dominance, and if this is consistent across species. Second, we examine how the ploidies of each species differ, and if the differences are congruent with phylogeny, or dominance. Third, we examine how much genome structural conservation remains between the species.

dN/dS

Genes contain coding regions that are transcribed, which in turn contain exons which are translated. Each three bases of an exon is referred to as a codon, each of which codes for a particular amino acid, although some amino acids can be coded for by multiple codons. Within the three bases of a codon there are sites which alter which amino acid is coded for, called non-synonymous sites, and sites that do not change the resulting amino acid, synonymous sites (Krause, 1995). Mutations occurring in the coding regions of a gene (exons) can be either synonymous or non-synonymous. The ratio of non-synonymous to synonymous sites is often called the dN/dS ratio, or omega (ω). Under a scenario of neutral evolution, the dN/dS ratio is expected to be 1 (Kimura, 1977; Goldman & Yang, 1994; Muse & Gaut, 1994). Deviations to either extreme of this scale can indicate that selection is happening at that site or locus, with a lower dN/dS suggesting purifying selection and a higher dN/dS suggesting positive selection (Kryazhimskiy & Plotkin, 2008; Jeffares et al., 2015). Analysis of dN/dS ratios requires high-quality coding sequences (CDS), combined into orthogroups, and accurately-aligned by codon. Several models are available, including

branch-specific models that allow investigation of variation in detected selection rates across multiple lineages in a predetermined phylogeny. Some models focus on detecting positive selection alone - aBSREL (Smith et al., 2015) or BUSTED (Murrell et al., 2015), and some allow inference of both purifying and positive selection - MEME (Murrell et al., 2012) and FEL (Kosakovsky Pond & Frost, 2005). aBSREL in particular looks at the number of sites where dNdS ratio is greater than 1, indicating the potential for positive selection or relaxation of purifying selection,

Through comparison of omega across multiple subgenomes and species we can look for regions of differential conservation, signalling that one subgenome or region is dominant over its homoeologous counterpart (Li et al., 2015; Yang et al., 2016; Cheng et al., 2018). If whole subgenome dominance is occurring we would see one subgenome with an overall lower omega than the other, with the dominant exhibiting low ratios and the non-dominant higher. If we observe high and low omega occurring in discrete genes on both subgenomes, but no overall difference between them, it would imply that whole subgenome dominance is not in effect and that the genome is undergoing a slower diploidisation process, if at all.

Relative rate tests

Comparing the rate of sequence change of each subgenome against homologous sequences from an outgroup can show us if there are any differences between the rate of molecular evolution between them, indicating if a locus or operational taxonomic unit (OTU) has been influenced by selection (Luikart et al., 2019; Tigano et al., 2020; Tichkule et al., 2021). One characteristic of asexual allopolyploids is a capacity for an increase in mutation rate due to transposable element activity, chromosome interactions and structural changes, and an increased point mutation rate due to subgenome redundancy (Selmecki et al., 2015; Vicient & Casacuberta, 2017). In light of this, we may expect to see different mutation rates across the MIG species. If one subgenome is significantly dominant over the other, we would also expect to see different mutation rates between subgenomes.

Differences in codon usage between subgenomes

The difference in codon usage between two subgenomes in an allopolyploid can inform us about the molecular evolution of the organism and its genome (Tian et al., 2020, 2022). Codon usage bias (CUB) is known to be influenced by several factors, primarily including overall GC content of the genome and factors influencing selection for a given codon, such as relative transcription or translation speed against synonyms and expression levels of the appropriate tRNA (Peden, 2000; Iriarte et al., 2021; Parvathy et al., 2022). Differences in

CUB between the subgenomes can suggest that they have been subjected to different evolutionary pressures, and through observing CUB across the MIG we can estimate how much, if anything, CUB diverges across monophyletic allopolyploid hybrids. Similarly to an increase in genome redundancy, polyploidisation brings more synonymous codons together in the hybrid giving a wider opportunity for selection, so we may expect to see CUB differences between species. CUB can also provide insights into the evolutionary history of the allopolyploid. If the two parental genomes had different CUB before hybridization, the resulting allopolyploid may exhibit a hybrid usage pattern, although this may be complicated by post-hybridization processes such as gene conversion or homogenization.

Ploidy profiling

Understanding the ploidy of an organism is important when interpreting the result of comparative and descriptive genome analysis, especially in determining the importance and activity of either subgenome (Schoenfelder & Fox, 2015; Meirmans et al., 2018). Subgenomes present in multiple copies will have higher levels of redundancy and tolerance to the increased mutational load that clonality drives, whereas those present in only one copy, such as the homoeologs of triploids, will not have such protection (Comai, 2005).

Different species of the same clade can exist in different ploidy states and the MIG is a good example. At least three levels of ploidy have been identified in the group: diploid, *M. floridensis;* triploid, *M. incognita;* tetraploid*, M. javanica* and *M. arenaria* (Eisenback & Triantaphyllou, 1991; Perry et al., 2009; Szitenberg et al., 2017; Winter et al., 2024). These ploidy differences can also exist within species, with different "races" predicted through cytology to exhibit different ploidy states (Eisenback & Triantaphyllou, 1991; Perry et al., 2009; Szitenberg et al., 2017; Winter et al., 2024). Notable examples from the MIG are *M. arenaria*, which has been observed as pentaploid (Blanc-Mathieu et al., 2017), and *M. incognita*, which exhibits diploids, albeit rarely (Triantaphyllou, 1981).

Methylation

Methylation is expected to combat genomic shock in allopolyploids by balancing gene dosage through the silencing of redundant genes (Shan et al., 2024), however methylation has not been succinctly detected in *Meloidogyne*. Detection of methylation in *M. javanica* could lead to discovery of methylation in the wider *Meloidogyne* genus, and if detected, comparison of rates between subgenomes could provide information as to how the species coped with genomic shock post hybridisation, as well as any potential subgenome

dominance interaction or fractionation controlled by methylation of non-dominant genes or regions.

Synteny

Synteny is the measure of conservation of homologous genes between two regions or chromosomes. A finer aspect of synteny is collinearity, which requires a conserved order alongside conserved sequence (Duran et al., 2009; Y. Wang et al., 2012). Observations of synteny are often informative in genomes resulting from whole genome duplications (WGD) as it enables visualisation of structural conservation between subgenomes or progenitor species (Tang et al., 2008; Chen et al., 2020; Cerca et al., 2022). We will investigate the levels of synteny between homoeologous subgenomes of species in the group to see how conserved synteny is across species. Fractionation is expected to diminish the amount of detectable synteny between genes or regions. As such, if a larger amount of synteny is found between one subgenome than the other, it would suggest dominance of that subgenome in both species.

3.1.5 Phylogeny

Phylogenies have been generated of the MIG and wider clade based on mitochondrial (García & Sánchez-Puerta, 2015; Szitenberg et al., 2017) and ribosomal loci (Szitenberg et al., 2017; Álvarez-Ortega et al., 2019). These show that the MIG are very closely related and share a common origin. These trees however do not address the phylogeny of these species at a subgenome level. In an allopolyploid, different genomic regions can have different phylogenetic histories. For example, either subgenome will have its own evolutionary history, independent of one another from speciation of the parent species up until hybridisation. This phylogenetic variability can also occur at a gene level, with homeologous conversion and gene fractionation leading to independent phylogenies for some genes.

Subgenome scale phylogenies have been generated of the *Meloidogyne incognita* group in the past (Figure 3.3), however in the absence of a phased assembly orthologs were assigned to subgenomes using clustering analysis and pairwise similarity. Among other features, Szitenberg et al. (2017) described two well defined, high bootstrap clusters, each containing a subgenome from all included MIG species, split into A or B, evidencing that the subgenomes are more similar and closely related to their conspecific subgenome in another species, than they are to the homoeologous subgenome of the same species, and implying

that the group originate from one or more hybridisation events between the same parent species. Interestingly, the outgroup used for this study, *M. enterolobii,* also shows evidence of allopolyploidy, albeit with different parental species to those of the MIG. With phased genome assemblies, we can now perform a far more specific subgenome scale phylogenomic analysis of the group, and in doing so position the subgenomes of *M. luci* within the framework of the MIG hybrids species using this depth of resolution for the first time.

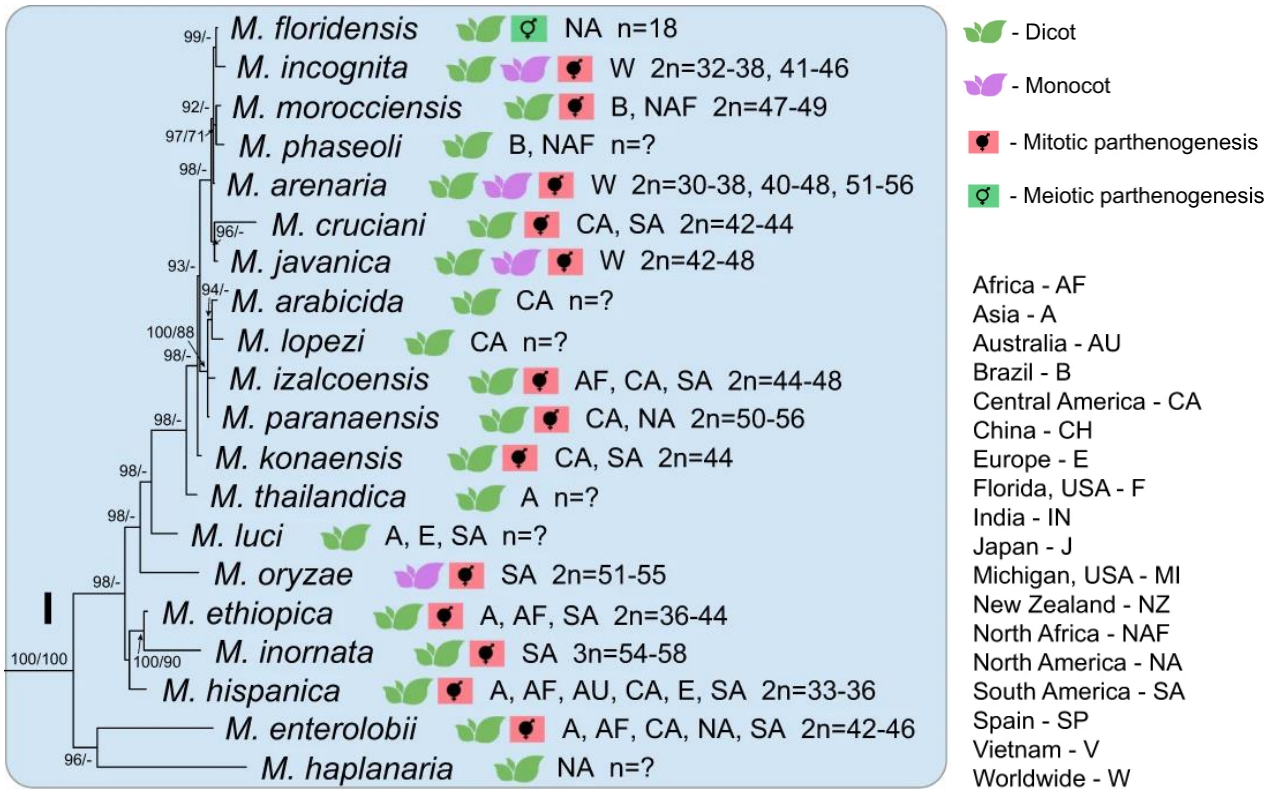


Figure 3.2: Phylogeny of clade I *Meloidogyne* species (Álvarez-Ortega et al., 2019). Inferred from 18S rRNA, ITS1 rRNA, D2-D3 expansion segments of 28S rRNA, COI gene and COII-16S rRNA sequence alignment. Icons display type of host and sexual system, followed by geographic location and predicted chromosome number.

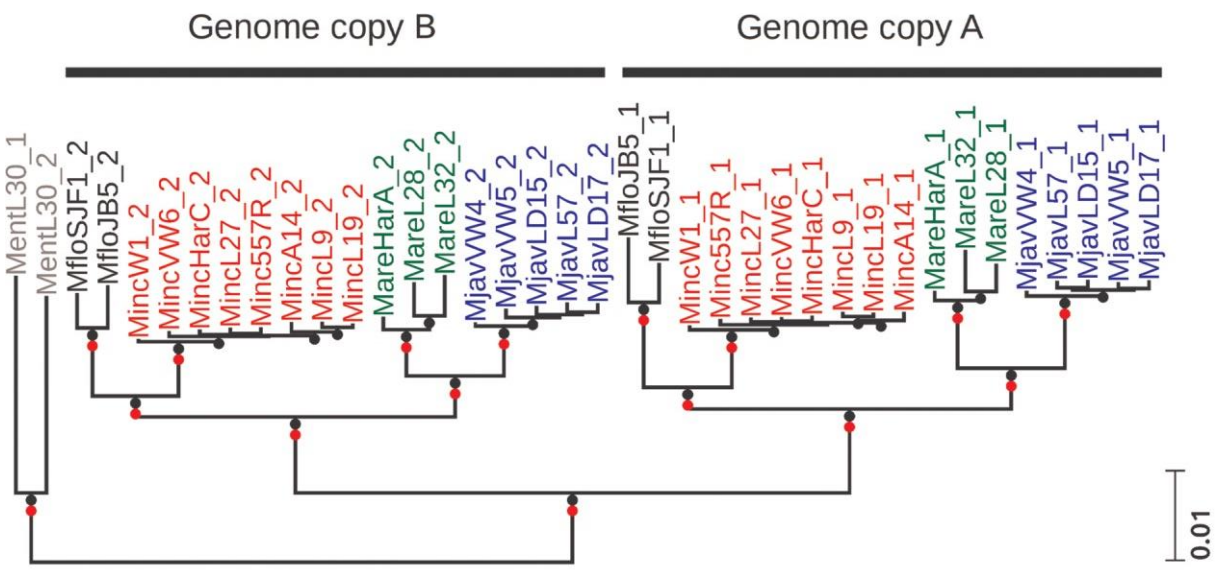


Figure 3.3: Subgenome phylogeny of the MIG (Szitenberg et al. 2017). The phylogeny of genes diverges into two clusters, each with a similar topology but split into genome copies, strongly indicative of the existence and preservation of divergent subgenomes within these species that share a homologous origin.

3.2 Methods

3.2.1 Reproducibility

In line with our aim of reproducibility, all data and scripts used within this work are available publicly online. Accessions for all data, including assemblies and sequence libraries, have been provided within the text at first mention of that dataset.

3.2.2 Acquisition of assemblies, annotations, and other data

We obtained genome assemblies of *Meloidogyne incognita* (MINJ2, ACC: GCA_014132215.1; Asamizu et al., 2020), *Meloidogyne arenaria* (MAA2, ACC: GCA_017562155.1), and *Meloidogyne luci* (MLUCI, ACC: GCA_902706615.1; Susic et al., 2019) from NCBI Genbank (Sayers et al., 2022), henceforth referred to as the contemporary assemblies. High accuracy short read libraries used to generate ploidy and genome profiles were obtained from Sequence Read Archive (ACC: Mi: SRR4242460, Ma: SRR4646457). CDS annotations of these species and their associated assemblies (Mi: ACC: GCA_003693645.1, Ma: ACC: GCA_003693565.1) generated by Szitenberg et al (2017)

were obtained from WormBase Parasite (Howe et al., 2017), along with gene predictions for *M. luci* (Danchin & Rancurel, 2023). The genome assembly (ACC: GCA_034785575.1) and annotations of *Meloidogyne javanica* were obtained from NCBI Genbank. CDS annotations for *M. javanica* were extracted from the main set using *AGAT: Another Genome Analysis Toolkit* (Dainat & Hereñú, 2020). CDS annotations of *Meloidogyne enterolobii* – an outgroup for phylogenetic analysis – were also obtained (Danchin et al., 2020; Koutsovoulos et al., 2020).

3.2.3 Ploidy profiling

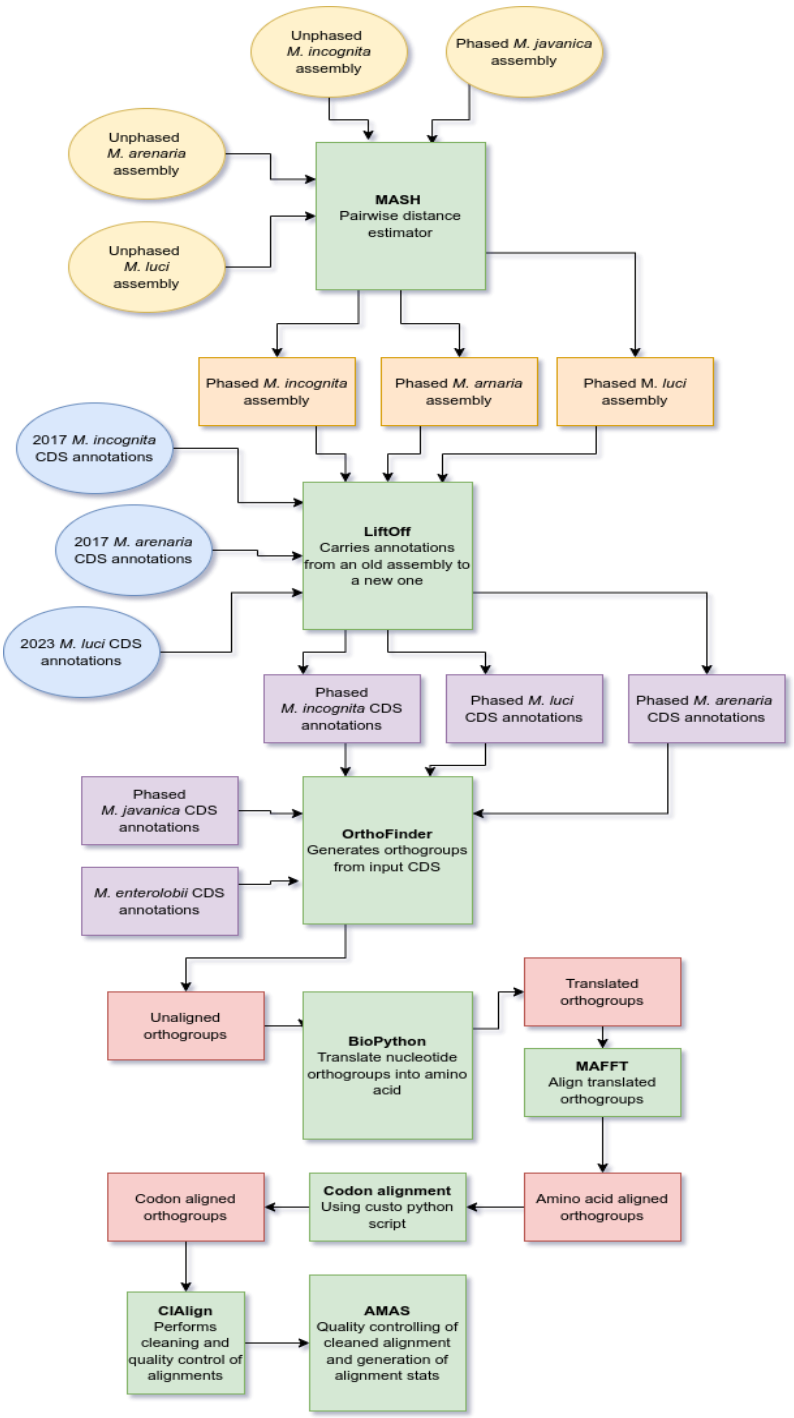
K-mer spectra and distributions for each high accuracy read library were generated using *KMC* with the parameters *-k21 -t28 -m1000 -ci1 -cs10000* (Kokot et al., 2017). The histograms from these distributions were then input to *smudgeplot*. Following ploidy determination with *smudgeplot*, each k-mer distribution histogram was input to *Genomescope2* with the appropriate ploidy setting as determined by *smudgeplot* (Ranallo-Benavidez et al., 2020).

3.2.4 Phasing of assemblies through alignment to subgenomes of *Meloidogyne javanica*

Genome assemblies of *M. incognita*, *M. arenaria*, and *M. luci* were assigned to corresponding scaffolds of the genome assembly of *M. javanica* using *MASH* (Ondov et al., 2016). The individual scaffolds from the contemporary assemblies were each queried against a *MASH* sketch of the *M. javanica* assembly. Each contemporary scaffold was then assigned to an *M. javanica* scaffold based on the best hit with the lowest *MASH* distance and significant p-value (<0.01). If the *M. javanica* scaffold hit was assigned to a subgenome, the corresponding contemporary scaffold was assigned accordingly. These pairings were assessed through local alignment in *BLAST* (Altschul et al., 1990).

3.2.5 Annotation of contemporary assemblies and subsequent phasing of CDS

M. incognita and *M. arenaria* assemblies were annotated by a lift-over approach, utilising the *LiftOff* package (Shumate & Salzberg, 2021). *LiftOff* was provided with the



unannotated contemporary assemblies, an older annotated assembly of the relevant species, and the corresponding CDS annotations. CDS annotations were carried over to the new assemblies and their sequences extracted based on the phase of the scaffold that they mapped to, producing a set of CDS for each subgenome for each species, with the exclusion of *M. enterolobii*. For *M. luci*, annotations were available and obtained for the contemporary assembly (Danchin & Rancurel, 2023) and phased based on phase assignment of the *M. javanica* scaffold onto which they were annotated. The disparity between the annotation methods deployed for each species may introduce some skew to the analysis, however generation and application of a consistent annotation method across all species included was beyond the scope of this investigation.

Figure 3.4: Pipeline used to generate quality controlled codon alignments from CDS of unphased *Meloidogyne* assemblies. Green indicates bioinformatic processes. Yellow indicates input assemblies. Orange indicates phased assemblies. Blue indicates unphased CDS annotations. Purple indicates phased CDS. Red indicates orthogroups.

3.2.6 Generation of multispecies and *M. javanica* specific orthogroups

Orthogroups - groups of homologous genes descended from a single common ancestor - were generated using *OrthoFinder2* (Emms & Kelly, 2018). *OrthoFinder2* was provided with CDS sequences from all OTUs - both subgenomes of *M. incognita*, *M. arenaria*, *M. luci*, and *M. javanica*. Only CDS located on phased scaffolds were included. As well as this we included CDS from an outgroup, *Meloidogyne enterolobii*, unphased and as a single OTU. Gene trees were generated for each orthogroup. For *M. javanica* specific orthogroups, only CDS from *M. enterolobii* and both subgenomes of *M. javanica* were included. In both cases, only the longest sequence for each OTU was retained, and shorter sequences were removed from the orthogroups. Counts of the number of orthogroups and their respective contents were performed (Supplementary Table 3.2 & 3.3) and only orthogroups containing sequences from all OTUs were carried forward. Gene trees were also inferred for all orthogroups passing these thresholds.

3.2.7 Codon alignment of orthogroups and quality control

Codon alignments of orthogroups were generated in the following way. First, nucleotide alignments were generated using *MAFFT* with default settings. Unaligned orthogroups were then translated from nucleotides into amino acid sequences using *Biopython* (Cock et al., 2009) and also aligned using *MAFFT* (Katoh et al., 2002) with default settings. Orthogroups were then aligned by codon using *Biopython*, incorporating aligned nucleotide orthogroups and corresponding translated and aligned amino acid orthogroups.

Codon alignments were then assessed and quality controlled in the following ways. All orthogroups were analysed using *CIAlign* (Tumescheit et al., 2022) with the *--visualise*, *--clean*, and *–interpret* modules, producing quality and content metrics, alignment visualisations, and cleaned versions of the alignments. Alignments were also assessed using *AMAS* (Borowiec, 2016) to generate summary statistics. Selected *AMAS* and *CIAlign* results were visualised together to detect correlations between metrics (Supplementary Figure 3). Alignments were then checked for the presence of stop codons and frame shifts using *Biopython* and custom scripts. Alignments containing premature stop codons were removed from the analysis.

3.2.8 Phylogenetic tree generation

Taxon-specific sequences in codon aligned orthogroups were programmatically combined end-to-end to generate a concatenated multi-gene alignment. This alignment was assessed for quality and generation of summary statistics using *CIAlign* with the *--visualise*, *--clean*, and *–interpret* modules enabled. The alignment was then input to *IQTREE* (Minh et al., 2020) using the -GTR nucleotide model (GTR; 1000 bootstraps; Kosiol et al., 2007) to produce a phylogenetic subgenome scale species tree. Trees were drawn with *toytree*, and *ete3* (Huerta-Cepas et al., 2016; Eaton, 2020).

3.2.9 Analysis of codon frequency and usage

To obtain species specific codon usage information for each taxon, sequences in codon aligned orthogroups were extracted and combined to produce frame-aligned OTU specific sequence sets. We generated statistics of CUB and frequency for each subgenome and outgroup using *codonW* (Peden, 2005), as well as a browser based solution (Stothard, 2000), which were in agreement (Supplementary Figure 3.1; Supplementary Table 3.1).

3.2.10 Detecting different rates of evolution between subgenomes of *M. javanica*

dN/dS

dN/dS ratio was calculated using *HyPhy* at the command line (Kosakovsky Pond et al., 2020). All orthogroup alignments that passed quality control were analysed using the aBSREL model in HyPhy (Smith et al., 2015; Kosakovsky Pond et al., 2020). Due to this model being branch specific all OTUs barring *M. enterolobii* were considered and set as foreground branches. Orthogroup specific phylogenetic gene trees were used for each group analysis in order to not skew results where the gene tree was incongruent with the species tree (Mendes & Hahn, 2016). Results were parsed, filtered for significance, and combined (Supplementary Table 3.2), and distributions of omega values for significant results plotted (Figure 3.8).

For loci with significant p-values (p = <0.05) for both subgenomes of *M. javanica* under the aBSREL model, the omega (ω) of both taxa was extracted and plotted alongside each other to display the distribution between the subgenomes for a given gene to visualise cases

where both subgenomes at a locus are being conserved, or alternatively where one subgenome is being conserved and the other has degenerated (Figure 3.9).

Relative rate

We compared the rate of sequence change of each OTU against homologous sequences from *M. enterolobii*. We performed quality control and checked for homogeneity of bases in the alignment (Table 3.4), then performed Tajima's relative rate test on the concatenated alignment. The test was applied in R using the *pegas* package (Paradis, 2010). Results were collated in a table (Table 3.5).

3.2.11 Methylation differences

Methylation was detected by re-basecalling the nanopore data (SRA: SRR23627521) using the *dna_r9.4.1_450bps_modbases_5mc_hac_prom.cfg* model to detect CpG modification. Data was base-called with the latest version of *guppy*, methylated CpG sites were identified using *nanopolish* (Loman et al., 2015). A call over 2 log likelihood is considered methylated and reliable. Calls over this threshold and detected in reads at a frequency of >40% were plotted along the length of each scaffold for all scaffolds (Supplementary Figure 3.2).

3.2.12 Multispecies synteny

A multispecies synteny analysis was performed using the *MCScan* module from the *JCVI* software (Wang et al., 2012; Tang et al., 2015) using the following parameters: C-score cutoff = 0.7, DIST = 20, N = 4. CDS annotations of all species were used to inform synteny detection. Subgenomes from each species were compared to their homologous subgenome in *M. javanica* to infer synteny (Figure 3.12, 3.12 & 3.13).

3.3 Results

Genome assemblies of *M. arenaria, M. incognita, M. javanica, M. enterolobii*, and *M. luci* were obtained and analysed using the *asmapp* workflow to determine assembly quality (Winter, 2022; Winter et al., 2024). Results are presented in Table 3.1. *BUSCO* scores for all assemblies were acceptable, ranging from 69.5% to 74.1%. *M. javanica* and *M. arenaria*

are expected to have similar genome sizes of around 280-300 Mbp based on previous studies, calculations from collapsed short read assemblies (Blanc-Mathieu et al., 2017; Szitenberg et al., 2017), and k-mer spectra (Figure 3.5). We observe similar lengths for *M. arenaria* and for *M. javanica* when accounting for genome length due to dual assembly. GC percentage is consistent across species at ~30% (Table 3.1).

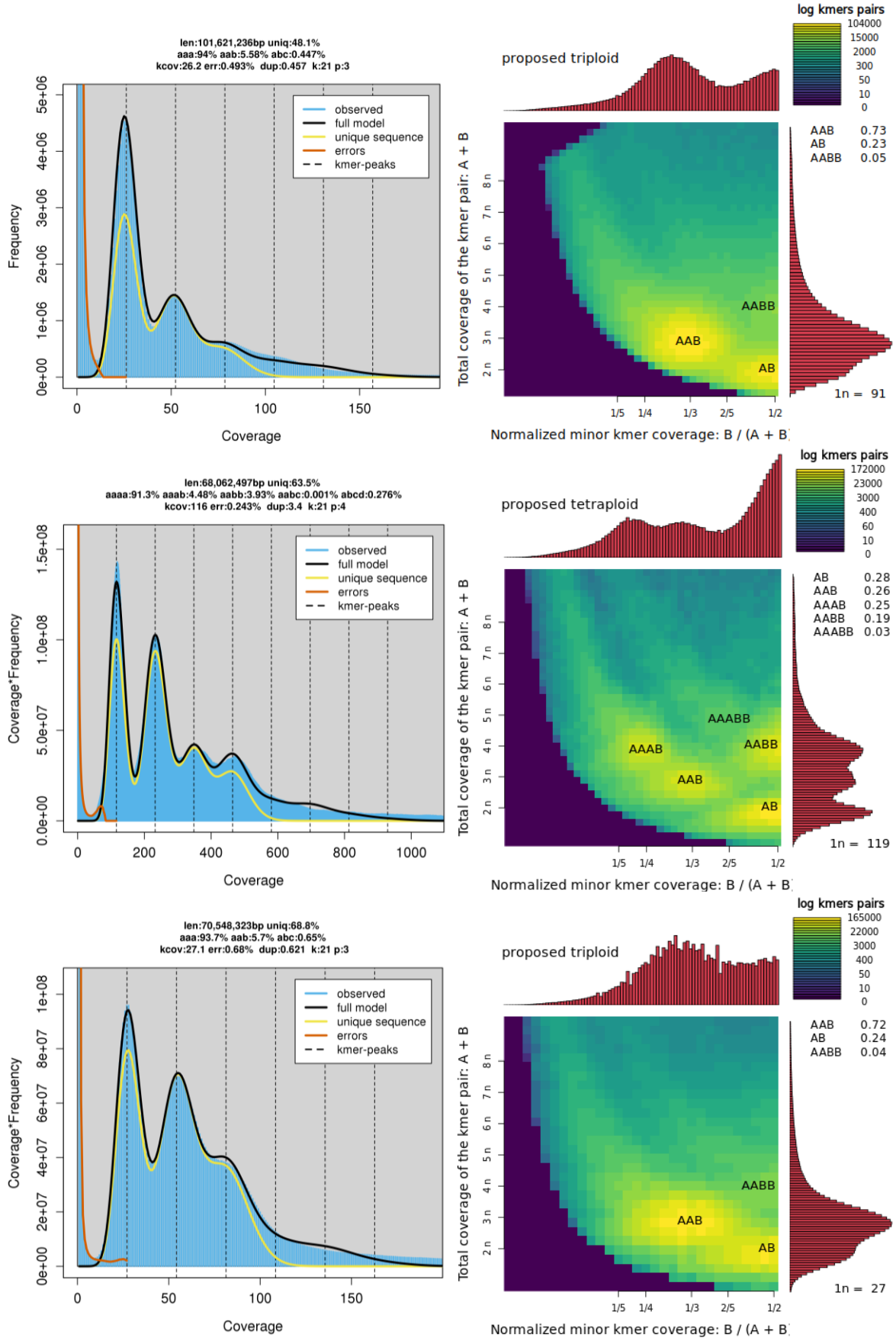
Table 3.1: Descriptive statistics of assemblies utilised in this study.

Species	Assembly length (Mbp)	BUSCO	GC (%)	Contigs	Year	Accession
M. arenaria	281.69	C:74.1% [S:16.1%, D:58.0%], F:10.2%, M:15.7%	30	1431	2021	GCA_017562155.1
M. javanica	150.54	C:69.5 [S:37.3%, D:32.2%], F:13.7%, M:16.8%	30.1	69	2024	GCA_0347 85575.1
M. incognita	193.15	C:71.7% [S:18.8%, D:52.9%], F:11.4%, M:16.9%	29.94	374	2020	GCA_014132215.1
M. luci	209.15	C:73.7% [S:14.9%, D:58.8%], F:11.0%, M:15.3%	30.17	327	2019	GCA_902706615.1
M. enterolobii	240.05	C:73.7% [S:13.3%, D:60.4%], F:10.6%, M:15.7%	30	4437	2021	GCA_9039 94135.1

3.3.1 Genome and ploidy profiling

The consensus between observed and modelled peaks in k-mer spectra show that ploidy selection was accurate and we produced models that accurately fit the k-mer spectra profile. We see different levels of ploidy across the species. Ploidy predictions are as follows: *M. luci* and *M. incognita,* triploid. *M. arenaria* and *M. javanica,* tetraploid.

We also see similarities between MIG species and differences from the more phylogenetically distant *M. luci.* Predicted haploid length for MIG species ranges from 68 Mbp to 75 Mbp. Predicted haploid length for *M. luci* is much higher, at ~102 Mbp. Unique k-mer percentages for MIG species are between 62.8% and 68.8%, however we observe a much lower percentage of unique k-mers in *M.luci*, at 48.1%. Duplication rates are between 0.402 and 0.621 for all species with the exception of *M. javanica,* for which we detect a duplication rate of 3.4.


len:101,621,236bp uniq:48.1%
aaa:94% aab:5.58% abc:0.447%
kcov:26.2 err:0.493% dup:0.457 k:21 p:3
observed
full model
unique sequence
errors
kmer-peaks
Frequency
Coverage
proposed triploid
log kmers pairs
104000
15000
2000
300
50
10
0
AAB 0.73
AB 0.23
AABB 0.05
AABB
AAB
AB
1n = 91
Total coverage of the kmer pair: A + B
Normalized minor kmer coverage: B / (A + B)
len:68,062,497bp uniq:63.5%
aaaa:91.3% aaab:4.48% aabb:3.93% aabc:0.001% abcd:0.276%
kcov:116 err:0.243% dup:3.4 k:21 p:4
Coverage*Frequency
proposed tetraploid
172000
23000
3000
400
60
10
0
AB 0.28
AAB 0.26
AAAB 0.25
AABB 0.19
AAABB 0.03
AAABB
AAAB
AABB
1n = 119
len:70,548,323bp uniq:68.8%
aaa:93.7% aab:5.7% abc:0.65%
kcov:27.1 err:0.68% dup:0.621 k:21 p:3
165000
22000
AAB 0.72
AB 0.24
AABB 0.04
1n = 27

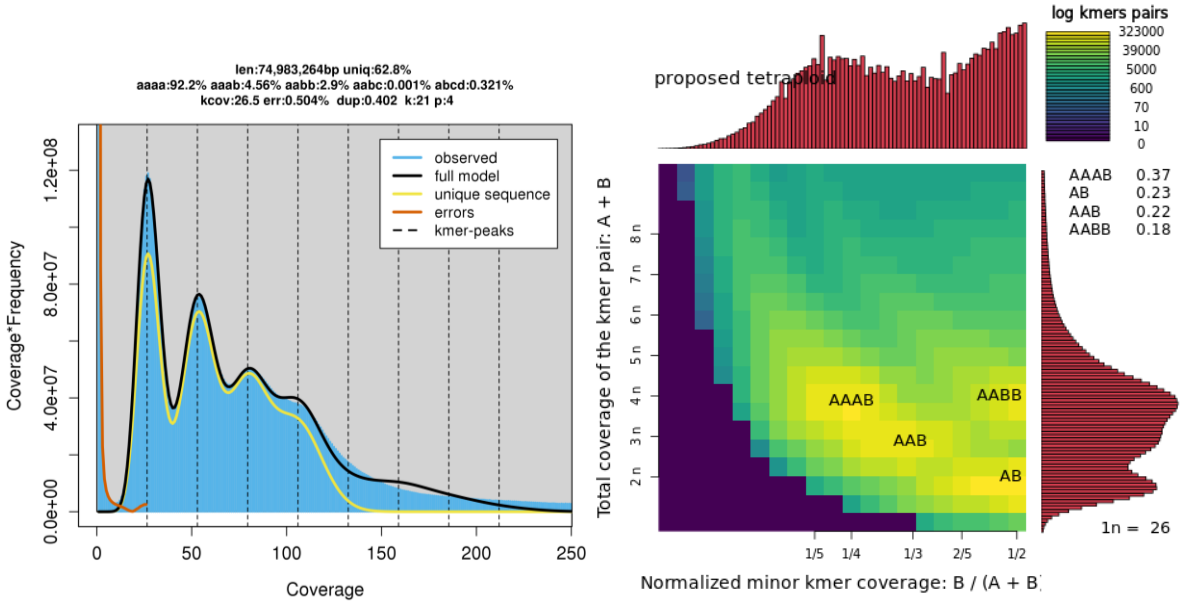


Figure 3.5: *Smudgeplot* and *Genomescope* results for species included in these analyses. Top to bottom: *M. luci*, *M. javanica, M. incognita, M. arenaria*.

3.3.2 Phasing of assemblies through alignment to orthologous subgenomes in *Meloidogyne javanica*

Winter et al. (2024; Chapter 2) phased scaffolds of an assembly of the *M. javanica* genome into subgenomes, assigned A and B. We then assigned contigs from the genome assemblies of *M. arenaria, M. incognita,* and *M. luci* to A and B subgenomes using *MASH* and *BLAST* similarity searches. Descriptive statistics of these assignments can be seen in Table 3.2. The majority of *M. incognita* and *M. luci* assemblies were assigned to a subgenome, at 78% and 76.87% respectively. Unfortunately, many contigs of the *M. arenaria* assembly did not hit with enough significance to be considered a match, and as a result 56.94% of that assembly was assigned to a subgenome. Due to only 87% of the *M. javanica* assembly being phased, the highest expectation of percentage of length phased for the other assemblies is also ~87%. In this context, we have phased 89.7%, 65.5%, and 88.36% of *M. incognita, M. arenaria,* and *M. luci,* respectively, out of the total amount possible using this method.

Table 3.2: Statistics of assembly contig assignment and phasing.

Species	*M. javanica*	*M. incognita*	*M. arenaria*	*M. luci*
Total contigs	33	374	1431	327
Contigs assigned	N/A	291	763	223
Percent of contigs assigned (%)	N/A	77.81	53.32	68.19
Phased to A (number and length)	11	97	260	81
Phased to B (number and length)	17	141	367	97
Length of contigs phased to A (bp)	58,935,666	53,631,533	56,803,706	71,458,573
Length of contigs phased to B (bp)	69,619,210	96,938,673	101,634,373	88,284,288
Percent of assembly phased (%)	87	78	56.94	76.87
Length of phased content (bp)	128,554,876	150,570,206	158,438079	159,742,861
A/B length ratio of phased contigs	1:1.18	1:1.81	1:1.77	1:1.23

3.3.3 Contemporary assembly annotation and phasing of CDS

Table 3.3: Statistics of CDS assignment and phasing.

Species	Total CDS	CDS lifted over	CDS assigned to phased contigs	Phased to A	Phased to B	Percent of CDS phased	A/B CDS ratio
M. javanica	35,749	NA	20,888	9412	11076	58.43	1:1.18
M. incognita	NA	21,507	17,301	6,031	11,270	80.44	1:1.87
M. arenaria	NA	26,692	16,175	5,178	10,997	60.59	1:2.12
M. luci	49,988	NA	39,635	17,316	22,319	79.29	1:1.29

We observe a similar number of genes assigned to subgenome B for MIG species, and roughly double for *M. luci*. The total number of CDS annotated on the assemblies is also double for *M. luci* than it is for MIG species. The number of CDS phased to subgenome A varies across all species which could be suggestive of dominance of subgenome B.

Ratios of CDS between subgenomes align roughly with ratios of subgenome length (1:1.18 - 1:1.87; Table 3.2) including for *M. luci* despite containing nearly double the CDS. The exception to this is *M. arenaria* with a higher skew of CDS (1:2.12) than subgenome length 1:1.77.

3.3.4 Orthogroup generation and alignment

OrthoFinder2 assigned 143,635 genes (93.7% of total from all species) to 19,177 orthogroups. 50% of all genes were in orthogroups with 10 or more genes (G50 = 10) and were contained in the largest 5291 orthogroups (O50 = 5291). There were 842 orthogroups with all species present and 0 of these consisted entirely of single-copy genes (Table 3.4). Full statistics of orthogroup generation can be found in Supplementary Tables 3.2.1 - 3.3.4.

Table 3.4: Overall orthogroup statistics

Metric	Count
Number of species	9
Number of genes	153,372
Number of genes in orthogroups	143,635
Number of unassigned genes	9,737
Percentage of genes in orthogroups	93.7
Percentage of unassigned genes	6.3
Number of orthogroups	19,177
Number of species-specific orthogroups	2,223
Number of genes in species-specific orthogroups	7,105
Percentage of genes in species-specific orthogroups	4.6
Mean orthogroup size	7.5
Median orthogroup size	6
G50 (assigned genes)	11
G50 (all genes)	10
O50 (assigned genes)	4,824
O50 (all genes)	5,291
Number of orthogroups with all species present	842
Number of single-copy orthogroups	0

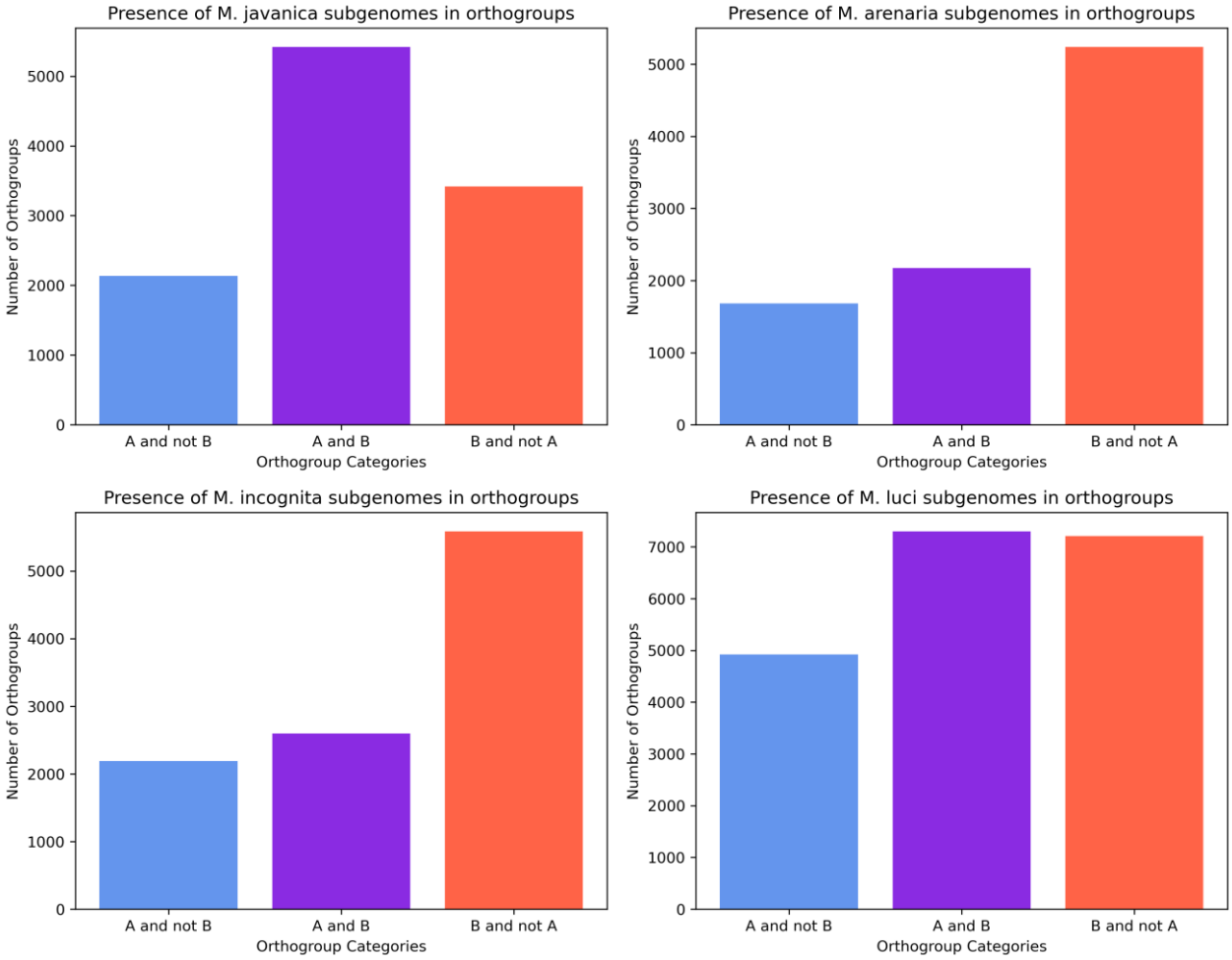


Figure 3.6: Histograms of subgenome presence or absence in orthogroups. The Y axis represents the number of orthogroups. The X axis represents the category: blue, subgenome A; purple, both subgenome A and subgenome B; red, subgenome B. Top left, *M. javanica*. Top right, *M. arenaria*. Bottom left, *M. incognita*. Bottom right, *M. luci*.

Not all orthogroups contain genes from both subgenomes for each species, and the rate at which genes from either subgenome are absent from orthogroups can inform us about the potential rate of gene fractionation or loss on a subgenome. Orthogroups containing genes from both subgenomes are the most common, although there are more orthogroups containing only subgenome B genes than both A and B for *M. luci*. There is a higher rate of orthogroups with only subgenome B copies of genes than subgenome A for all OTUs, and we see a higher number of subgenome B genes assigned to orthogroups overall (Supplementary Table 3.3) suggesting that subgenome A has undergone a higher rate of gene fractionation than subgenome B (Figure 3.6). These results are indicative of subgenome B being the most dominant subgenome, although these differences could be

explained by the relative greater total length of phased subgenome B contigs, leading to a higher number of input CDS overall (Table 3.3).

3.3.5 Generation and quality control of codon alignments

842 orthogroups were aligned by both amino acids and nucleotides. Of these orthogroup alignments, 205 successfully aligned by codon, with the remaining alignments containing ambiguous or missing data (Tumescheit et al. 2022). 22 orthogroups were found to include sequences that contained stop codons. These orthogroups were removed from the analysis, bringing the total number of orthogroups retained to 183 (Supplementary 3.2). These alignments were parsed with *CIAlign* and *AMAS* to generate descriptive statistics and observe the interaction between metrics (Supplementary 3.3.1 and 3.3.2).

Table 3.5: Whole genome and subgenome scale homogeneity of bases. Whole counts are from whole genome including non-coding and intergenic regions. Subgenome counts are from CDS regions only.

Species	State	A (%)	T (%)	G (%)	C (%)	AT (%)	GC (%)	Skew (%)
Meloidogyne javanica	Whole	36.67	38.33	16.67	8.33	75.03	24.97	3.005:1
	A	35	28.33	18.33	18.33	63.33	36.66	1.73:1
	B	18.33	51.67	18.33	11.67	70	30	2.33:1
Meloidogyne incognita	Whole	42.5	22.5	15	20	65	35	1.86:1
	A	28.33	31.67	21.67	18.33	60	40	1.5:1
	B	26.67	38.33	18.33	16.67	65	35	1.86:1
Meloidogyne arenaria	Whole	35	36.25	12.5	16.25	71.25	28.75	2.48:1
	A	26.67	33.33	23.33	16.67	60	40	1.5:1
	B	30	30	21.67	18.33	60	40	1.5:1
Meloidogyne luci	Whole	20	43.75	18.75	17.5	63.75	36.25	1.76:1
	A	55	21.67	10	13.33	76.67	23.33	3.29:1
	B	21.67	40	20	18.33	61.67	38.33	1.61:1
Meloidogyne enterolobii	Whole	33.75	33.75	15	17.5	67.50	32.50	2.08:1

We calculated the ratio of bases for CDS in each orthogroup alignment reaching this stage of analysis and checked for homogeneity of bases to ensure appropriateness to

undergo tests of selection (Figure 3.7; Table 3.4). CDS from triploid species have an AT/GC ratio of 1.76-1.86:1 and an AT percentage of between 63.75% and 65%. CDS from tetraploid species have a ratio of 2.48-3.01:1 and an AT percentage of between 71.25% and 75.03 percent.

In *M. incognita* and *M. javanica*, CDS on subgenome B have a higher percentage of AT bases, whereas in *M. luci* CDS on subgenome A have the higher percentage, with subgenome A CDS of *M. luci* exhibiting the highest percentage of adenine of all bases of all subgenomes from all species. The ratio of AT/GC is the same in CDS from both subgenomes of *M. arenaria*.

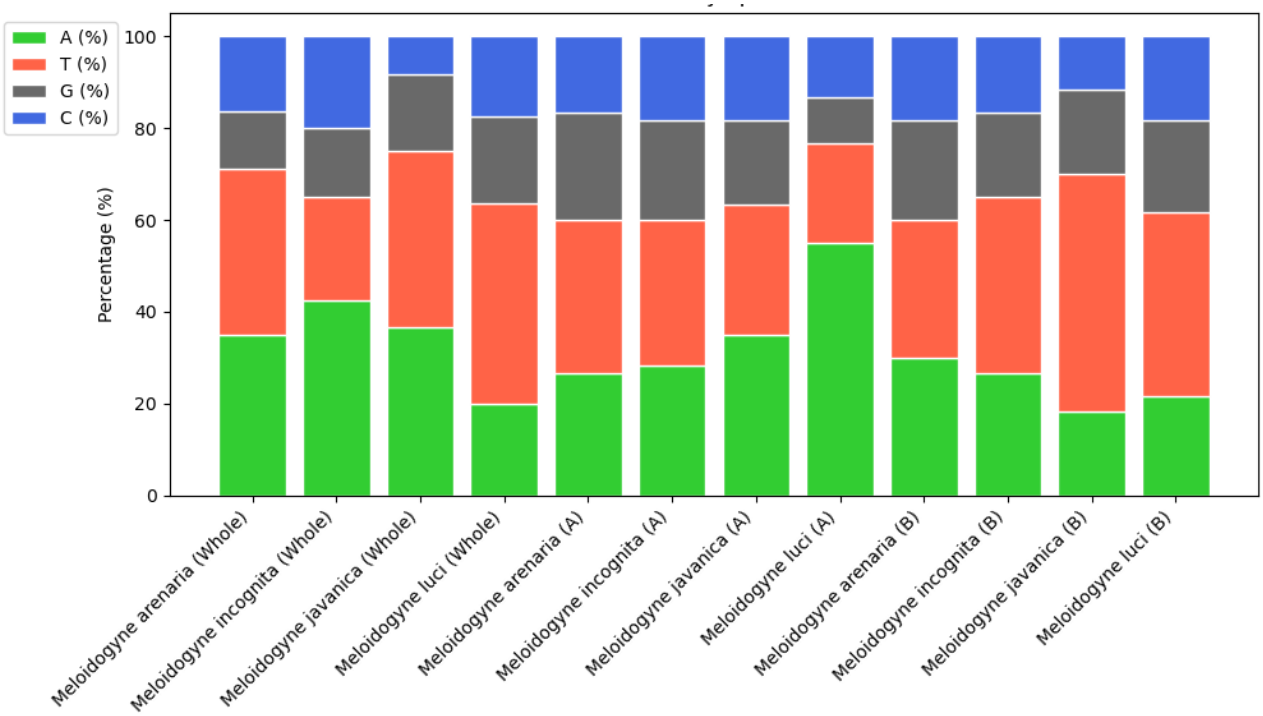


Figure 3.7: Distribution of base usage in whole genomes and subgenomes. Divided into the whole genome and each subgenome for every species analysed. Adenosine, green. Thymine, red. Guanine, grey. Cytosine, blue.

3.3.6 Phylogenomics

Structure of the subgenome phylogeny shows that subgenomes from all included taxa are more similar and more closely related to their contemporary subgenome in other species (*M. javanica* A to *M. luci* A) than they are to the homoeologous subgenome (*M. javanica* A to *M. javanica* B) they share the organism with (Figure 3.8).

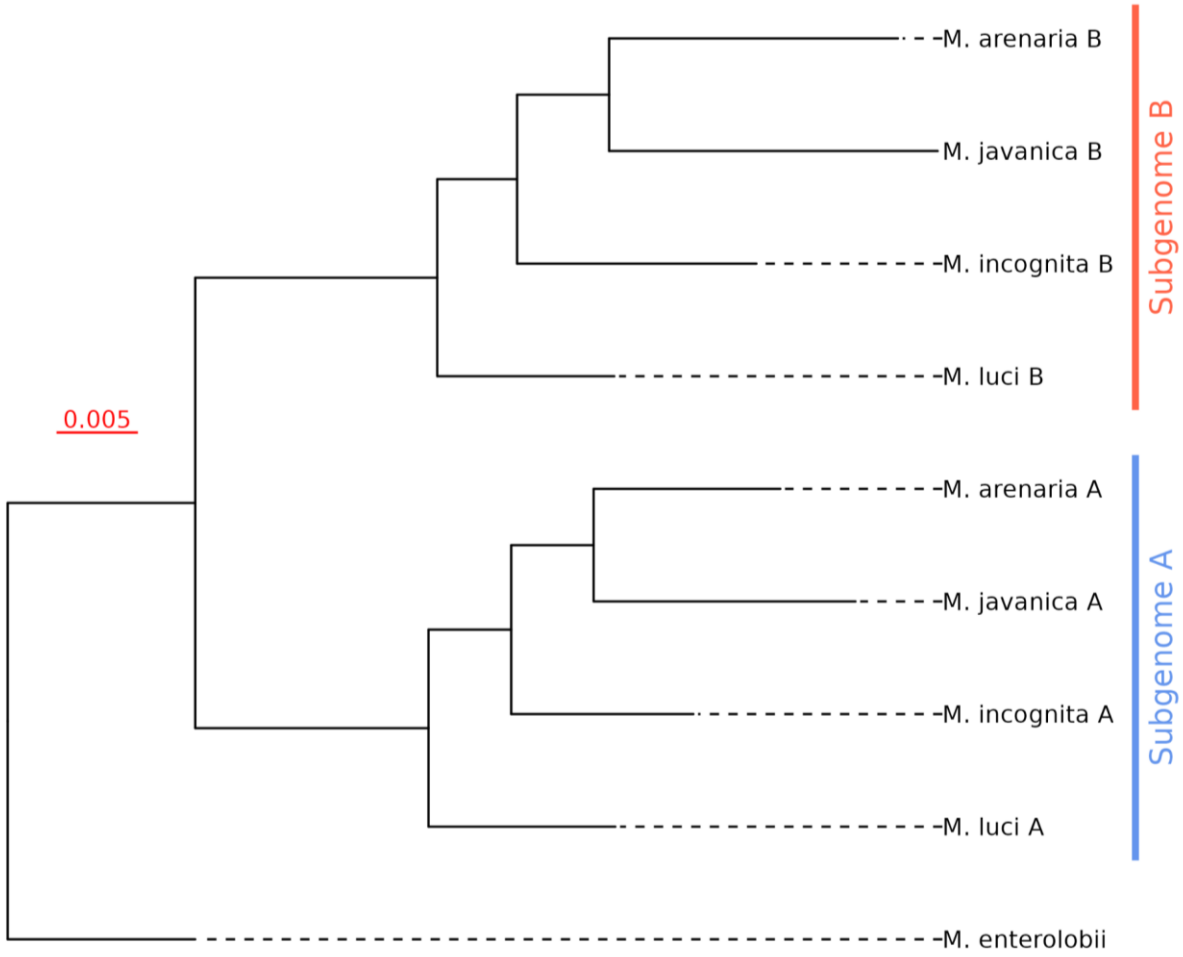


Figure 3.8: Phylogenomic subgenome-specific species tree. Built from concatenated phased and aligned CDS orthologs from subgenomes of *M. javanica, M incognita, M. arenaria, M. luci,* and an outgroup, *M. enterolobii*.

3.3.7 Codon usage and frequency

We compared codons across CDS for all species and observed no difference in usage or frequency between species or subgenomes (Supplementary Figure 3.1; Supplementary Table 3.1).

3.3.8 Detection of selection in *Meloidogyne javanica*

We performed tests of selection using the aBSREL branch-specific model on all orthogroups containing sequences from all 9 OTUs to detect differences in selection pressures between homeologous genes across subgenomes, and to attempt to infer if one subgenome was dominant over the other based on overall amounts of positive or negative selection.

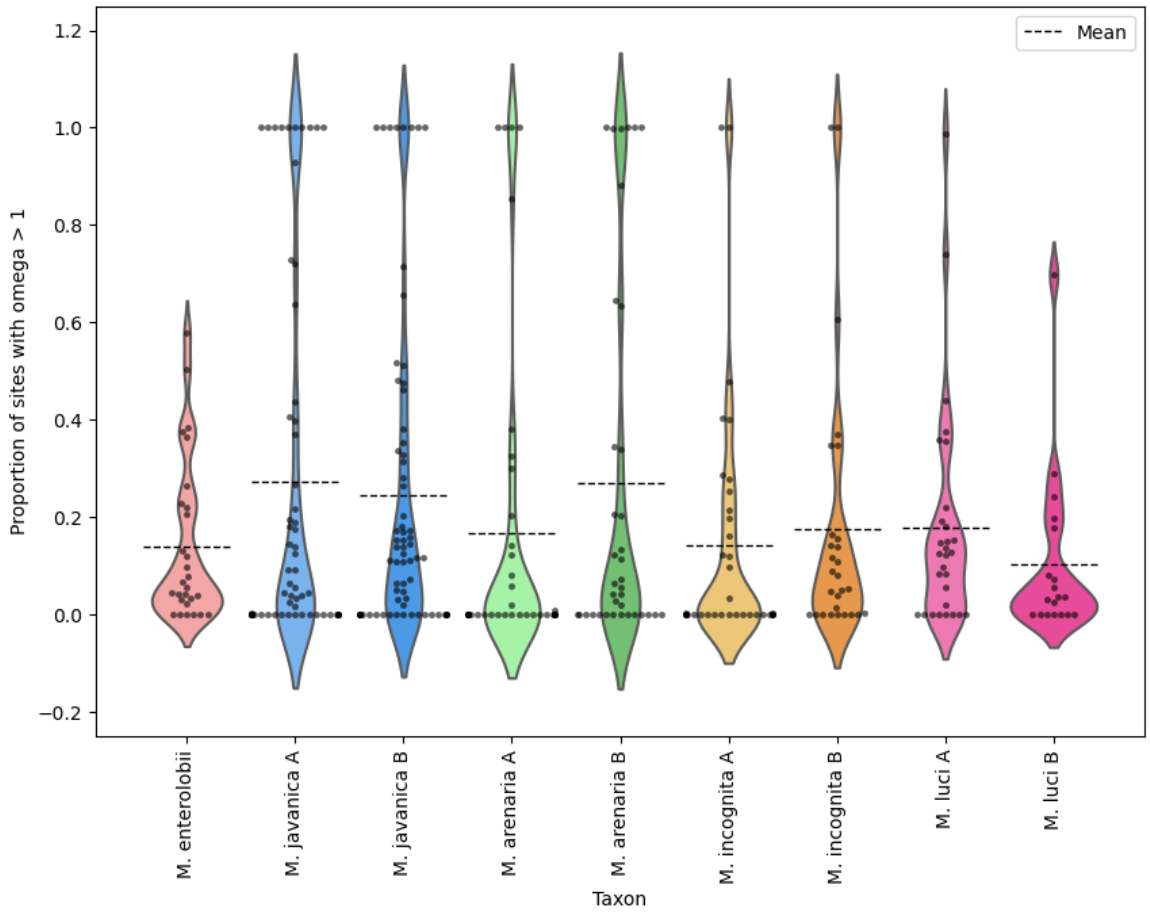


Figure 3.9: Violin plot of orthogroups containing sites with omega (ω) greater than 1 for all OTUs analysed. Colour indicates species, with lighter shade indicating subgenome A and darker shade indicating subgenome B.

Subgenome level

On a subgenome level, we observe a higher mean omega value in subgenome B sequences for all species except for *M.luci* (Figure 3.9). The distribution of omega values for each OTU was plotted to investigate the spread and frequency of omega for each (Figure 3.10). We see that in *M. javanica, M. incognita,* and *M. luci,* there is a higher frequency of genes on subgenome A with high omega than subgenome B although this effect is very minimal in the latter two species. In *M. arenaria* we see a higher frequency of genes of subgenome B with high omega. We see a higher frequency of genes under negative selection in subgenome A for triploid species (*M. incognita* and *M. luci*), and a higher frequency in subgenome B for the tetraploid species (*M. arenaria* and *M. javanica*). We also observe a greater proportion of genes with lower omega in triploid species than in tetraploid species (Figure 3.10).

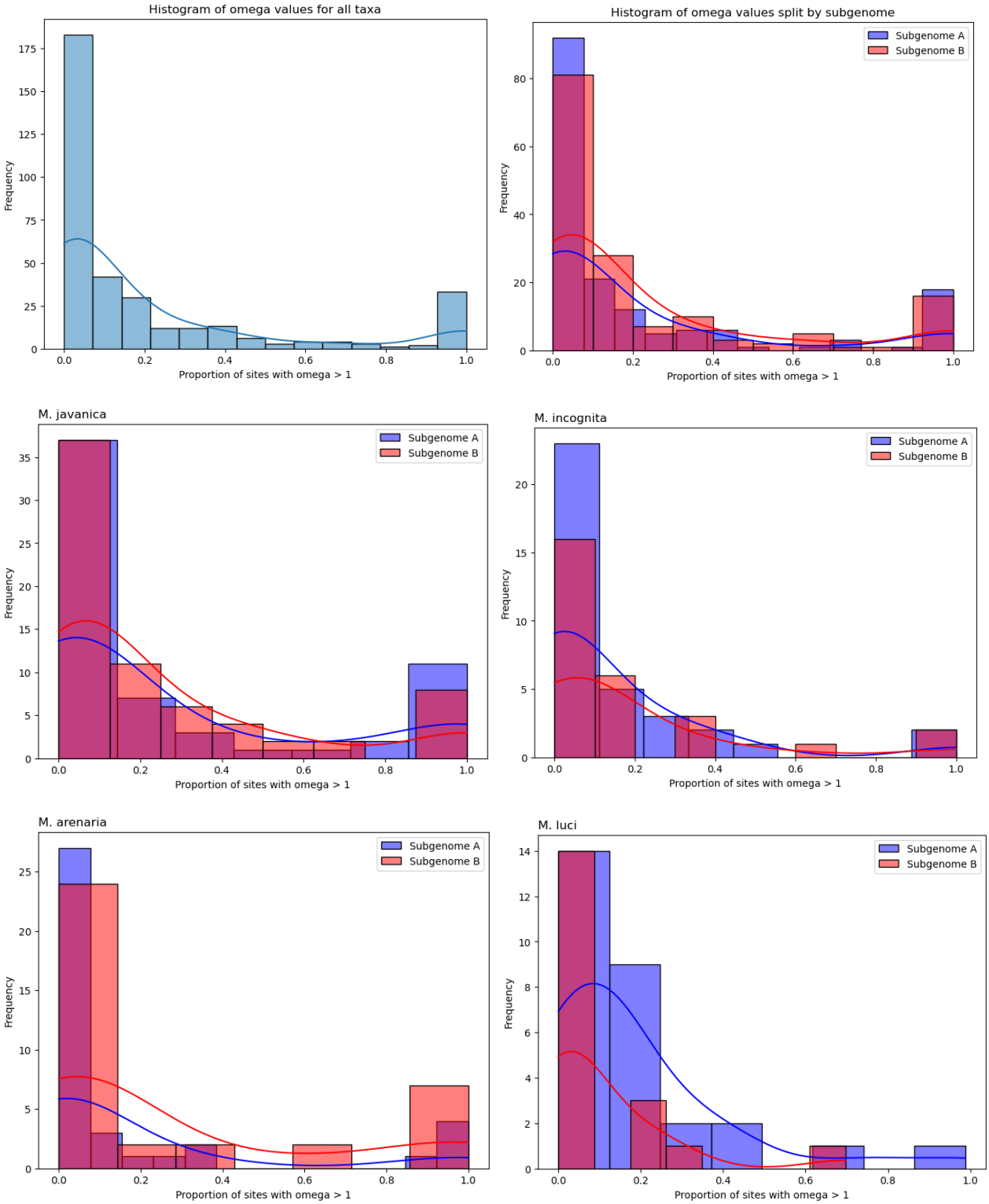


Figure 3.10: Distribution of orthogroups with sites showing omega (ω) > 1. Curves represent average trend of the bars. **Top left,** Proportion of sites with **ω** > 1 for all taxa. **Top right,** Proportion of sites with **ω** > 1 for all species in analysis divided into subgenome A and subgenome B. **Middle left,** Proportion of sites with **ω** > 1 for *M. javanica* divided into subgenome A and subgenome B. **Middle right,** Proportion of sites with **ω** > 1 for *M. arenaria* divided into subgenome A and subgenome B. **Bottom left,** Proportion of sites with **ω** > 1

for *M. incognita* divided into subgenome A and subgenome B. **Bottom right,** Proportion of sites with **ω** > 1 for *M. luci* divided into subgenome A and subgenome B.

These results combined suggest that subgenome A is dominant over subgenome B, resulting in subgenome A experiencing a higher conservative selection pressure, enabling occasional neofunctionalisation of subgenome B genes.

Locus level

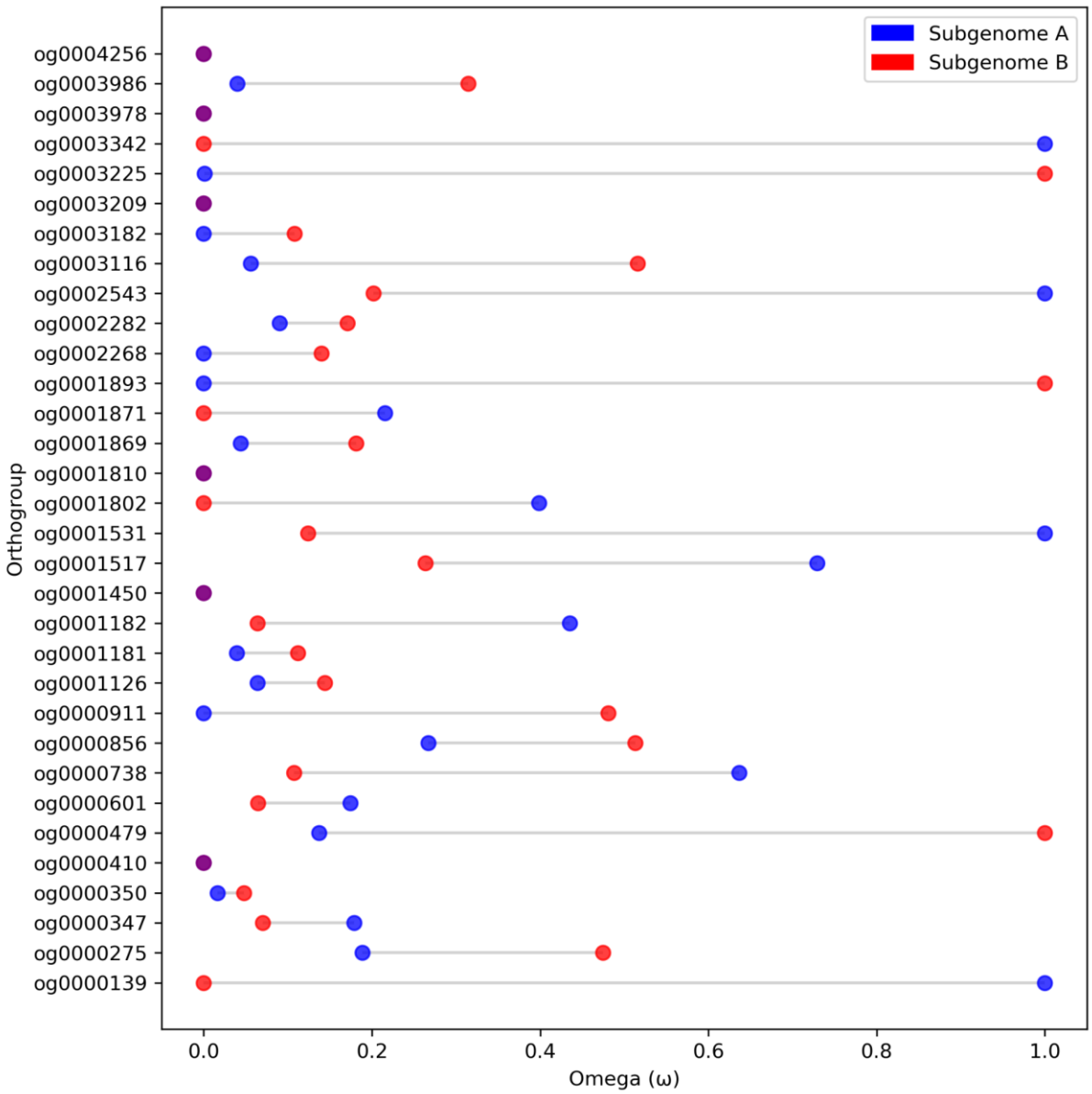


Figure 3.11: Distribution of significant homoeologous dNdS scores for orthogroups containing CDS from subgenome A and subgenome B of *M. javanica*. X axis shows

omega, where scores of **ω** > 1 are scaled to 1. Only orthogroups where genes from both subgenomes were found to be significant are included.

From the results of analysis with *HyPhy*, we selected orthogroups where genes from both subgenomes of *M. javanica* exhibited significant omega values (32 orthogroups) and plotted them for comparison (Figure 3.11). We see a large disparity in omega in 7 of them. These results suggest that genes on opposing subgenomes experience opposite selection pressures, where one copy is conserved under negative selection and the other falls under either neutral or positive selection. We observe 16 instances where subgenome A is under negative selection while subgenome B is not - although only 5 of these cases have an omega of >0.5 for subgenome B - and only 9 where subgenome B has the lower omega value. Even in these cases the omega value tends to be higher than those of genes where subgenome A is under negative selection. All together these results suggest that subgenome A is being conserved more favourable than subgenome B, indicating that subgenome A is dominant in *M. javanica*.

3.3.9 Relative rates

Of all species tested, only *M. luci* did not exhibit a significant difference in relative rate of nucleotide substitution between subgenomes. All of the MIG species tested showed a significant difference in relative rate between subgenomes.

Table 3.6: Results of Tajima's relative rate test. Generated using *pegas* in *R.*

Species	***M. javanica***	***M. arenaria***	***M. luci***	***M. incognita***
Chi	18.8048	46.76605	0.02471042	33.01541
p-value	<0.001	<0.001	0.875	<0.001

3.3.10 Methylation

Raw Oxford Nanopore reads were only available for *M. javanica* which limited our methylation analysis to the subgenomes of this species. We observe a methylation rate of ~3% across both subgenomes. When visualised, this methylation does not appear to be randomly distributed across the genome (Supplementary Figure 2), and some calls are at a frequency of 100%.

3.3.11 Synteny

We detected and compared the amount of conserved synteny between each subgenome of *M. javanica*, *M. arenaria, M. incognita,* and *M. luci.* We find a considerable amount of synteny between all species and *M. javanica, s*o much so between some chromosomes and syntenic assembly scaffolds that it could be used as reference to pseudoscaffold the opposing contigs. *M. luci* contains almost chromosome scale runs of synteny with *M. javanica* in both subgenomes. Something to consider is that the higher rate of fragmentation of the *M. arenaria* and *M. incognita* assemblies could be impacting detection of syntenic blocks due to programmatic constraints on the length of syntenic blocks, increasing false negatives and missing syntenic locations. Conversely, higher contiguity of the *M. luci* assembly could be facilitating the joining of syntenic blocks. We are unable to infer chromosome fragmentation of other species relative to *M. javanica* due to lack of continuity in the contemporary genome assemblies. This analysis likely underestimates the amount of synteny between species due to annotations of the contemporary species being based on a lift over approach from older annotations.

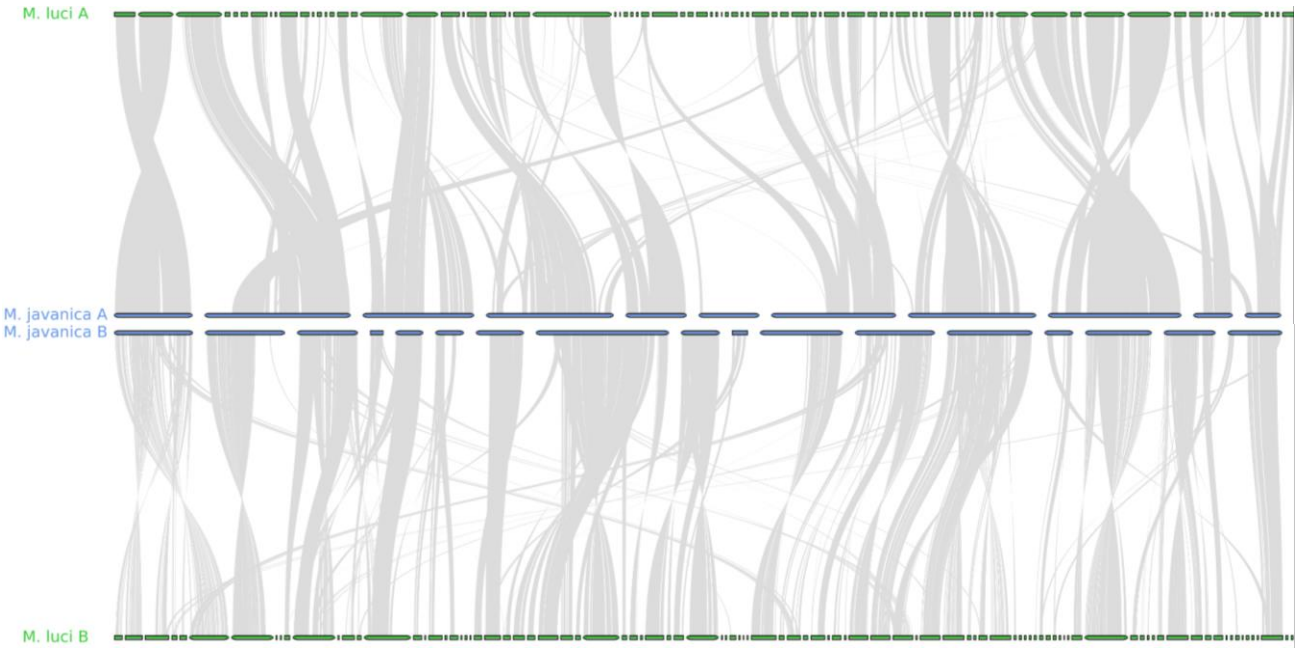


Figure 3.12: Plots showing synteny and collinearity between subgenomes of *M. javanica* and *M. luci.*

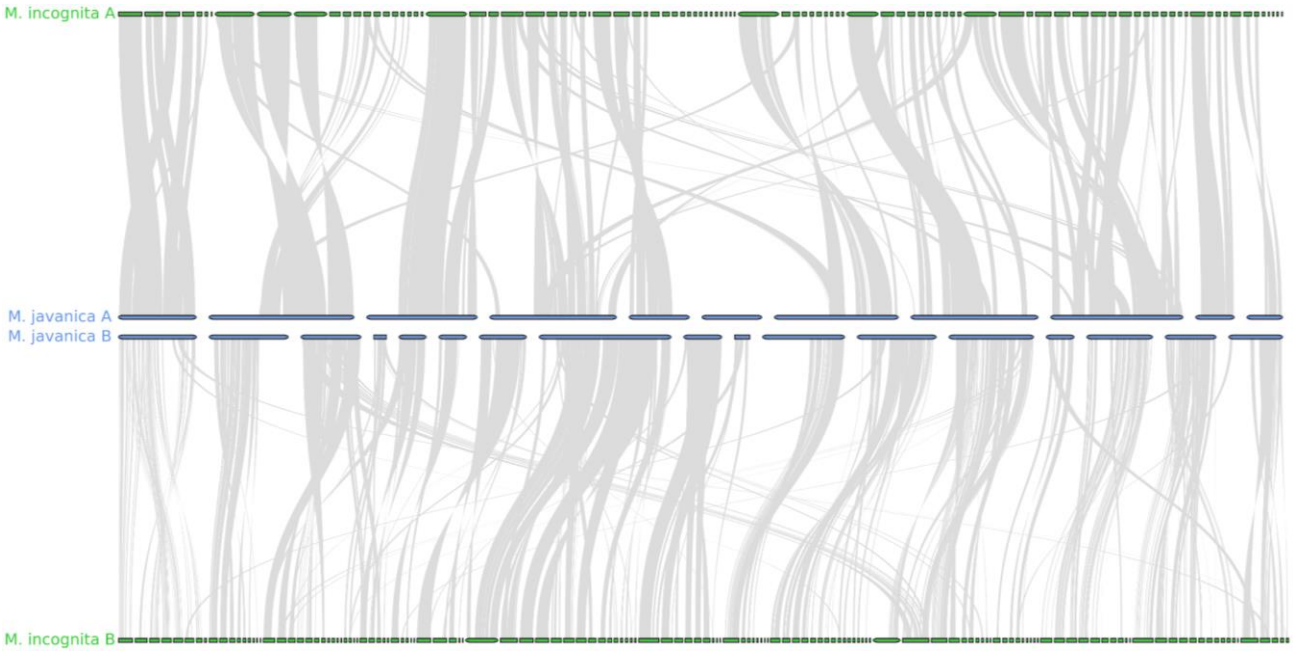


Figure 3.13: Plots showing synteny and collinearity between subgenomes of *M. javanica* and *M. incognita*.

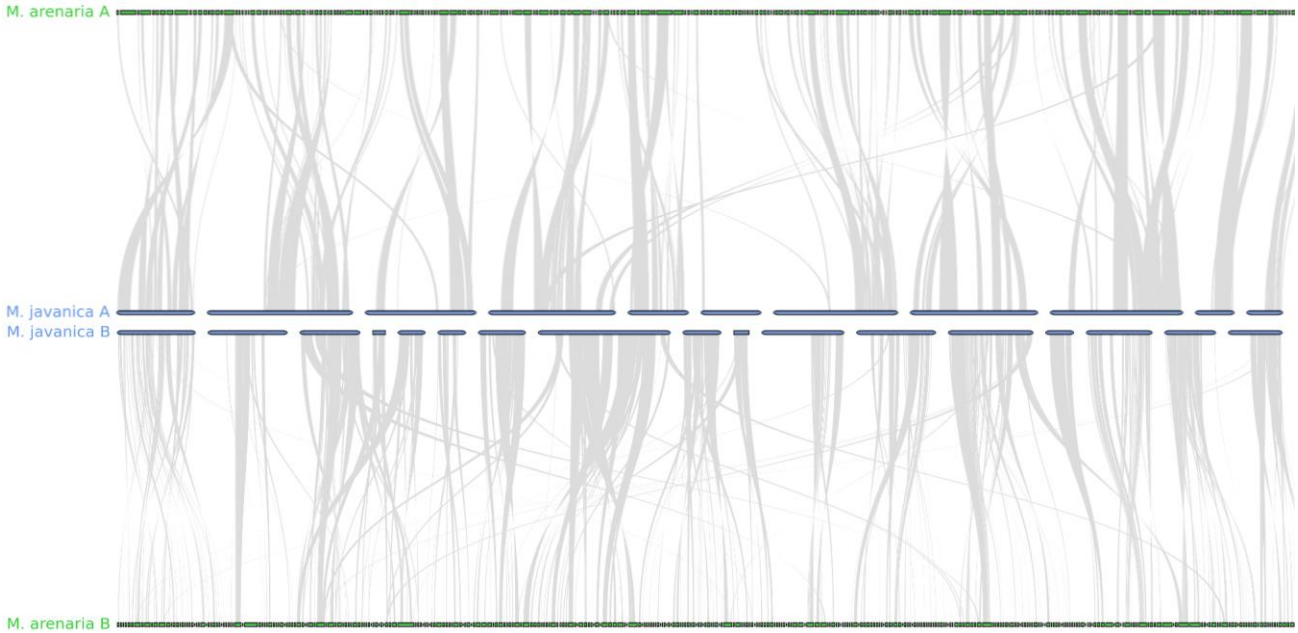


Figure 3.14: Plots showing synteny and collinearity between subgenomes of *M. javanica* and *M. arenaria.*

3.4 Discussion

We set out to compare the subgenomes of several RKN species: *M. javanica, M. arenaria, M. incognita,* and *M. luci.* We successfully phased assemblies of the latter three species using a recent phased, chromosome-scale assembly of *M. javanica* as a reference (Winter et al., 2024), and subsequently generated descriptive statistics of the assemblies, subgenomes, and respective CDS counts. Overall we consider this method of lifting over and phasing CDS successful and believe that it has provided enough phased CDS for each species to enable further analysis to be representative. Using these phased CDS sets, we generated groups of orthologous genes and investigated for signals of selection or differences in mutation rates and effects. Alongside this, we assessed ploidy of all species, synteny and codon usage frequency between subgenomes across species, performed a phylogenomic analysis of all subgenomes, and attempted to detect methylation rates of subgenomes of *M. javanica*.

3.4.1 Subgenome dominance

The ratio of subgenome length to amount of phased CDS per OTU is roughly congruent (~1:1.18), which is to be expected if CDS are evenly distributed throughout the genome and no single subgenome has lost significantly more genes than the other (Table 3.2 & 3.3). We observed a greater amount of orthogroups containing only CDS from subgenome B for all species (Figure 3.6), which could suggest that subgenome A has undergone more gene loss than subgenome B, although the imbalance in the number of CDS from each subgenome, when input to *OrthoFinder2,* is likely driving this difference. This is important to consider when attempting to infer subgenome dominance from the presence of CDS, which based on Figure 3.6 suggests dominance of subgenome B. There may also have been a skewed ratio of total CDS between the ancestors of the hybrid, though this is difficult to determine. Performing further analysis on only orthogroups containing CDS from all OTUs will remove this skew, however it will affect counts of orthogroups containing CDS from only one subgenome, which potentially hold signals of imbalance gene loss.

Rates of omega across genes from orthogroups where all subgenomes and species are represented suggest that subgenome A is dominant in the triploid species, *M. incognita* and *M. luci*, with a higher frequency of very low omega genes suggestive of purifying selection. Dominance is more balanced in the tetraploid species, *M. javanica* and *M. arenaria* (Figure 3.10), although these effects are slight, with both tetraploid species showing

a greater amount of high omega genes than the triploid species tested. A more equal division of loci with lower omega between subgenomes could indicate that dominance is more balanced in the tetraploids, and may explain why these species have not reverted to triploidy like *M. incognita* and *M. luci*.

We present this effect of dominance at a locus level scale in *M. javanica*, where we see that several genes across subgenomes display contrasting omega, seemingly under different selection pressures (Figure 3.11), evidencing the effect of localised subgenome dominance. This localised dominance appears to be balanced between subgenomes, leading to no significant detectable subgenome-scale trend.

All MIG species, bar one, showed significant differences in relative rates of substitution between their subgenomes (Table 2.5), highly suggestive of one subgenome being dominant over the other, with higher omega detectable on the other. In contrast *M. luci* does not show a significant difference in relative rate between subgenomes. The difference of relative rate between the MIG and *M. luci* (Table 3.6) could be explained by a different response to genomic shock in their progenitors following hybridisation, with the genome of *M. luci* stabilising earlier as a triploid, and subgenome dominance being asserted.

Overall, we find that no subgenome is significantly dominant over the other, likely facilitated by alternating locus-level dominance, with a balanced number of genes on either subgenome experiencing low omega whilst their homoeologs experience high omega. This has led to retention of both subgenomes in all species tested. Reversions to triploidy in some species could therefore be explained through homologous recombination accumulating coding genes on one copy of a subgenome, facilitating the degeneration of the other, which would lead to the AAB genome complements exhibited by *M. incognita* and *M. luci*. Analysis and detection of these phenomena can be complex, and in attempting to explore the genomes of these species from several other perspectives, our investigation of subgenome dominance is shallow compared to the depth required to conclusively determine the dynamics of this system. We have however laid the groundwork for more in depth and targeted analysis, as well as developed workflows and bioinformatic methods to aid and enable future research.

3.4.2 Phylogeny

We find that the subgenomes of the species in this study are more similar to their orthologous subgenome in sister species than they are to the homoeologous subgenome in the same cell, supporting phylogenies produced by Szitenberg et al (2017) and Winter (2020). These species share a common ancestor that was the hybrid progeny of two ancient *Meloidogyne* species.

Although this pattern is expected in the MIG species based on previous studies, *M. luci* also follows this pattern. In this work, subgenomes of *M. luci* were phased and placed in a subgenome scale phylogeny for the first time, revealing that *M. luci* subgenomes are paraphyletic, following the same topology and evolutionary history as the MIG - each subgenome more similar to its homologs than its homoeologs - but branching from the lineage more basally that the coalescence point of the MIG species (Figure 3.8).

This suggests a congruent hybrid origin followed by divergence into the progenitors of the MIG and *M. luci*, with *M. luci* diverging from the common ancestor earlier than the MIG speciated. A shared origin but earlier speciation could also be supported by our findings of few genes with high omega and no significant difference between the relative rate of mutation between *M. luci* subgenomes, indicative of a restored genome stability. It could be hypothesised that dominance of subgenome A was established early in *M. luci*, with fractionation occurring on subgenome B, triggering speciation, and inducing copy loss and the return to triploidy.

Future phylogenomic studies of this clade could include other apomictic species close in phylogeny to *M. luci* such as *M. oryzae, M. konaensis, M. ethiopica*, and possibly *M. haplanaria*, phased into subgenomes, and attempt to detect a similar pattern.

3.4.3 Ploidy differences

We determined ploidy states of species included in this study and presented them together for comparison for the first time (Figure 3.5) and observed differences in ploidy levels concordant with those previously inferred by past groups (Blanc-Mathieu et al., 2017; Szitenberg et al., 2017; Susič et al., 2020). Regarding differences in ploidy between species analysed here, phylogenomic analysis strongly suggests a shared origin resulting from hybridisation of the same parent species. However, we are unable to tell for certain from these results whether this was a single hybridisation event that gave rise to the ancestor of

the MIG and *M. luci* or if two separate hybridisation events occurred that resulted in the ancestor of *M. luci*, and the ancestor of the MIG separately. Despite this, we would suggest the more likely option of a single event.

The paraphyly of triploids containing homologous subgenomes suggests that the triploid species *M. incognita* and *M. luci* have both undergone copy loss and triploidisation independently, whilst *M. javanica* and *M. arenaria* have maintained tetraploidy. There are several mechanisms through which this disparity of ploidy could arise. It could be that purifying selection acting in equilibrium on both subgenomes at a local level in the tetraploids has offset the diploidisation process and enabled maintenance of polyploidy, whereas in the triploid species dominance settled with a greater bias toward one subgenome, accelerating copy loss (Comai, 2005; Bomblies et al., 2016). In any case, gene fractionation and diploidisation are the likely reasons behind the triploidy seen in *M. incognita* and *M. luci*.

An alternative explanation for triploidy in these species could be the possibility of reproducing sexually with rarely produced males, which have been identified in apomictic MIG species, although there is no record of fertilisation being observed despite successful insemination (Papadopoulou & Traintaphyllou, 1982). If possible, sexual reproduction would facilitate recombination and linkage of dominant genes and regions on the dominant subgenome, accelerating the rate of degeneration of the non-dominant subgenome. Rare sexual events in some species could also explain the anomalous sexual system and genome architecture of *Meloidogyne floridensis,* an outlier in the MIG by being automictic and predicted as either diploid or triploid (Lunt et al., 2014; Szitenberg et al., 2017; Jaron et al., 2021).

3.4.4 Methylation

The presence of methylation in the genomes of *Meloidogyne* has not been robustly confirmed before, and methylation in nematodes in general is thought to be rare, if existent at all (Gao et al., 2018; Pratx et al., 2018). We detect a genome-wide rate of 3% methylated CpG sites. This is low compared to the values found in taxa known conclusively to exhibit DNA methylation (Liu et al., 2020; Engelhardt et al., 2022), however, if rare a high rate would not be expected. There is also the possibility that this amount is within the range of potential false positives, although this 3% all pass the threshold of what is generally considered a true call (log likelihood >2), and methylation of reads covering some sites was detected at a frequency of 100%.

DNA methylation is a common mechanism for genomes undergoing WGD to manage genomic shock, with increased and potentially deleterious gene dosage controlled through silencing of genes through methylation (Keller & Yi, 2014; Song & Chen, 2015; de Tomás & Vicient, 2023). A supposed lack of considerable methylation in *Meloidogyne* raises questions about how, in the supposed absence of methylation, *Meloidogyne* species coped with genomic shock post-hybridisation. One possibility is the upregulation of piRNAs – PIWI-interacting RNAs thatprevent the movement of transposable elements – are able to assist in managing genomic shock (Czech et al. 2018; Loubalova et al. 2023).

3.4.5 Codon usage and frequency

The consistency we observe in codon usage could reflect genomic or biochemical constraints within the group, such as a limited tRNA repertoire or strong conservation of translational pathways, preventing divergence of codon usage following speciation. Similarity of CUB across subgenomes of the MIG and *M. luci* is also indicative of the homologous origin of their subgenomes.

3.4.6 Synteny

We find from the amount of synteny observed between subgenomes across species that our method of assigning phase to contigs of unphased assemblies was successful (Figure 3.12, 3.13 & 3.14). Synteny between subgenomes across species appears to be considerably conserved for both subgenomes, indicating that large structural rearrangements are rare between species, and suggesting that retention of genes is occurring in collinear blocks rather than at an individual gene level. It is however important to note that the lack of contiguity of some of the assemblies used in synteny analysis could be shrouding some structural rearrangements from our view, and that repetition of this analysis with more contiguous or even chromosome-scale genome assemblies would, if not reveal aspects that are unable to be detected here, provide a more confident estimation of the amount of retained collinearity.

3.4.7 *Meloidogyne luci*

Based on the results presented here, we would suggest that *M. luci* originated from the same hybridisation event as the MIG, but speciated earlier than the ancestor of the MIG

from their hybrid progenitor. Non-significant relative rate of substitution between subgenomes and the lower number of genes exhibiting high dN/dS omega in *M. luci* would suggest that subgenomes have stabilised faster than those of the MIG, which still display evidence of balancing selection and fractionation. Despite phylogenetic separation from the MIG and *M. javanica,* subgenomes of *M. luci* still share a large amount of synteny and collinearity with homologous subgenomes of *M. javanica,* and presumably the wider MIG (Figure 3.12, 3.12, & 3.13).

3.4.8 Benchmarking Universal Single Copy Orthologs

Similarly to previous studies (Blanc-Mathieu et al., 2017; Szitenberg et al., 2017; Koutsovoulos et al., 2020; Susič et al., 2020; Winter et al., 2024), we observe a limited score of complete and fragmented BUSCO genes across the species analysed. The number of complete BUSCOs are similar across species, further establishing that BUSCO databases are not properly representative of core genes in clade I of *Meloidogyne*, and that a BUSCO score of 70-75% can be considered representative of a complete assembly of these species by current standards, however few explanations have been given for why this is the case. The most likely reason is that the BUSCO databases for Nematoda and Eukaryota are not wholly representative of *Meloidogyne*, and that *Meloidogyne* has lost some of its BUSCO genes since divergence from species better represented in the databases.

The ratio of singleton to duplicate BUSCOs between *M. javanica* and the other species likely differs due to the *M. javanica* assembly being a dual assembly, containing collapsed haploid representations of both subgenomes (Winter et al., 2024). The rates of singleton to duplicated BUSCOs across all assemblies is indicative of gene loss at some of the alleles of these loci.

3.4.9 Future work

There are several analyses that we aimed to perform in this comparison but were ultimately unable to.

A comprehensive analysis of repeats, repeat content and placement, and assessment of transposable element (TE) activity could be very informative to the investigation of subgenome evolution across this group. TEs are thought to help cope with genomic shock, as well as facilitating adaptation and gene fractionation through revived TE activity and

translocation post hybridisation (Vicient & Casacuberta, 2017; de Tomás & Vicient, 2023). Another line of inquiry could be identification of unique k-mer spectra in ancestor specific repeat regions, then searching for these k-mers in the other subgenome to detect homeologous recombination.

A thorough analysis of indels across subgenomes of all species could help inform the rates and spatial density of gene loss, as well as informing of potential TE activity. Winter et al (2024) observed a high prevalence of indels between subgenomic copies (A1/A2, B1/B2), and although some indels can be inferred from synteny analysis, it is unfortunate that we cannot expand this here to detect the same patterns between homologous pairs

Although we have assessed genome wide methylation rates in *M. javanica*, more detailed analysis of methylation in *Meloidogyne* is a necessity. This can include comparison of methylation at a CDS level rather than genome-wide, as well as comparing these rates between subgenomes, and also rates between homoeologous scaffold pairs. If methylation is indeed present, due to its implications in the management of genomic shock post hybridisation, we would expect to see differing levels between homeologous CDS, and even between homeologous pairs. Ultimately, sequencing these species with the most recent Oxford Nanopore Technology sequencing chemistry, calling methylated bases with the most current algorithms, and performing bisulfite sequencing in tandem, is the only way to decisively validate that methylation is present in the genomes of these species, as well as confirming the presence or absence of DNA methylase machinery in the assemblies

Since this work was completed, assemblies of *M. incognita* and *M. arenaria* were published that look to be more complete and of a higher quality than those used in this analysis (Dai et al., 2023). Future analysis could include these assemblies either independently or alongside the assemblies used here in order to produce more resolute results, particularly in the case of synteny. In addition, an increased number of representative samples would increase confidence in phylogenomic analysis through increased robustness of the tree.

3.5.0 Conclusion

The major findings of this chapter are as follows. Synteny is conserved throughout the MIG and *M. luci*, which combined with the phylogenomic position of *M. luci* strongly suggests

homologous ancestry of both subgenomes, meaning that *M luci,* and presumably sister species such as *M. thailandica* and *M. arabicida*, most likely share the same parentage as the MIG.

We find enough evidence of CpG methylation present in *M. javanica* to warrant further and deeper investigation of the possible presence and adaptive potential of methylation in the MIG.

Our analysis suggests that subgenome dominance interactions are at play in the genomes of the allopolyploid clade I *Meloidogyne* species, and that subgenome A is likely the overall more dominant subgenome. However, the preservation of tetraploidy versus triploidy in these closely related species suggests that different pressures and routes of fractionation have been taken by each, experiencing dominance and subsequent copy loss at different strengths and speeds.

Overall, this group provides challenges to genomic inference with its complexity, while at the same time providing an excellent system with which to investigate the evolution of genomes post hybridisation, the mechanisms and effects of subgenome dominance in asexuals, and the evolutionary forces driving and resulting from these processes.

Chapter 4: Detecting candidates for gain-of-virulence mutations in *Meloidogyne javanica*

Michael R Winter[1]*, Adam P Taranto[2], Alison Coomer Blundell[2], David H Lunt[1], Valerie M Williamson[2], and Shahid Siddique[3]

[1] School of Natural Sciences, University of Hull, Kingston-upon-Hull, UK.

[2] Department of Plant Pathology, University of California Davis, Davis CA, USA.

[3] Department of Entomology and Nematology, University of California Davis, Davis CA, USA

Correspondence: mrmrwinter@gmail.com

4.0 Abstract

The Javan root knot nematode, *Meloidogyne javanica* significantly affects crop yields and agricultural productivity in infected farmland. Several control measures are available, one of which is the use of genetically resistant plant cultivars containing the *Mi-1* gene, on which *M. javanica* is unable to successfully infect and reproduce. Strains of *M. javanica* able to evade this engineered resistance have been identified (e.g. VW5), increasing the threat posed by this species. Previous attempts to characterise a gain-of-virulence mechanism have been limited due to a lack of modern genomic resources for *M. javanica* and the wider group, and an inability to assess genomic differences between lineages in the context of allopolyploidy. Here we utilise a recent chromosome-scale, phased genome assembly of *M. javanica* to investigate and characterise a region thought to be involved in gain-of-virulence and evasion of *Mi-1*; the *Cg-1* region. In doing so, we discovered a large deletion on one of four genomic copies of *M. javanica* VW5, spanning over ~650 Kbp and encompassing the *Cg-1* region. We describe this region and the transcriptionally active genes present within it, discovering that some transcripts heavily implicated in host-pathogen interactions are entirely deleted from VW5, suggesting a situation of adaptation through gene loss, whereby loss of a secreted effector could prevent the host from recognising infection and activating defensive resistance mechanisms.

4.1 Introduction

4.1.1 Root-knot nematodes and *Meloidogyne javanica*

Root-knot nematodes (RKN) are a group of over 90 species of microscopic roundworms that infect the roots of flowering plants, causing drastically reduced host fitness and reduced yield of affected agricultural crops (Perry et al., 2009; Ralmi, 2016; Khan et al., 2023). Of the many RKN species, some of the most damaging species belong to the *Meloidogyne incognita* group (MIG), including *Meloidogyne incognita, Meloidogyne arenaria, and* the focal species of this study *Meloidogyne javanica* (Bird, et al., 2009; Jones, A. Haegeman, et al., 2013).

The Javan root-knot nematode, *Meloidogyne javanica*, can be found all around the world, but is particularly widespread in Africa, Southern Asia, and Australia (Ralmi et al., 2016). *Meloidogyne javanica* infects hundreds of plant species but its primary hosts are *Solanum lycopersicum* (tomato), *Capsicum annuum* (bell pepper), *Solanum melongena* (aubergine), and *Solanum tuberosum* (potato) (Oka et al., 2012; Moosavi, 2015; Tariq-Khan et al., 2020).

MIG species reproduce through apomixis and exhibit allopolyploid genomes of differing ploidies (Blanc-Mathieu et al., 2017; Szitenberg et al., 2017; Winter et al., 2024). A recent genome assembly of the *M. javanica* lineage VW4 showed how the genome of *M. javanica* consists of two subgenomes, each inherited from ancestors that hybridised to generate the allopolyploid species. It was also shown that *M. javanica* is tetraploid, with two copies of each subgenome (Winter et al., 2023). Synteny between these subgenomes is generally consistent, with the exception of some regions, although the karyotype of each subgenome appears to differ, with subgenome A containing some chromosomes that pair with multiple chromosomes of subgenome B, suggesting chromosomal fission events.

4.1.2 Secreted effectors and virulence genes

RKN infect the roots of host plants and induce giant cell (GC) formation - enlarged, hypertrophied cells - on which the RKN feeds (Perry et al., 2009; Jones et al., 2013). The generation of GCs causes extensive galling of the host roots and leads to increased susceptibility to infection from other organisms, and effects ranging from reduced yield to death of the host (Favery et al., 2016). Successful invasion of the host and successive

formation of GCs is facilitated by the secretion of effector proteins by the RKN (Jagdale et al., 2021).

Parasitic effector proteins are enzymes produced in the organism and used to facilitate multiple stages of infection, from initial invasion of the host, to upregulating molecular machinery to increase the nutritional yield available to the parasite from the host (Dalio et al., 2018; Jagdale et al., 2021). Primarily secreted from oesophageal glands, effector proteins in RKN have been implicated with several functions, including plant cell wall degradation and the formation and maintenance of the GC (Favery et al., 2020).

4.1.3 Genetic control measures

Historic treatment and prevention methods for RKN infection have revolved around the use of pesticides or other exogenous methods, however as evidence of the detrimental effect of these products on both the environment and human health continues to build, a different approach is required (Feyisa, 2021). One such approach is the development of genetically modified (GM) host plants. In these GM cultivars, resistance to RKNs is innate, reducing or removing the need for exogenous intervention. The exact mechanism of GM resistance differs depending on the species of RKN, the host plant taxa, and on the mechanism of infection (Williamson, 1998; Perry et al., 2009).

4.1.4 *Mi-1* mediated engineered genetic resistance

Description and role in conferring resistance

Mi-1 is a gene introgressed into cultivated tomato variants that confers resistance against RKN infection (Williamson, 1998; Nombela et al., 2003). This method is particularly effective at enabling resistance to three of the MIG species; *M. incognita, M. javanica, and M. arenaria*. The *Mi-1* gene is now present in many commercially and publicly available strains of tomato.

Mi-1 is a group of genes collectively referred to as "*Mi-1"*, that includes several homologous genes from *Mi-1* to *Mi-7*. Of these genes, *Mi-1.2* is the only one able to encode a functional resistance gene, and we refer to *Mi-1.2* as *Mi-1* throughout this chapter (Seah et al., 2007).

Although such GM hosts are generally resistant to MIG RKN infection, lineages of several MIG species have been found to be able to evade *Mi-1* mediated resistance, and *M. javanica* is no exception. Some such lineages appear to have evolved this evasion adaptation in the wild, without exposure to *Mi-1* resistant hosts (Eddaoudi et al., 1997; Tzortzakakis et al., 2005), however some virulent lineages have emerged in a laboratory setting through selection on populations able to infect *Mi-1* resistant strains (Castagnone-Sereno et al., 1993; Noling, 2000; Joseph et al., 2016). One lineage of *Meloidogyne javanica* capable of *Mi-1* evasion is the VW5 strain (Milligan et al., 1998; Gleason, 2003).

Significance of *Mi-1* evasive strains in infection management

There is evidence of *Mi-1* evasive strains in the wild. This reduces the efficacy of *Mi-1* and increases the risk of infection in engineered host plant strains. Over the long term, rising global temperatures could potentially enable the range of highly impactful MIG species to increase, enabling the colonisation of agricultural land previously not considered threatened. Increasing global temperatures could facilitate the reproduction and survival of MIG species in areas that were historically too cold for them to thrive (Dutta & Phani, 2023). This does not only include areas latitudinally distinct from current ranges, but also those at higher altitudes than those previously able to support the species. Combined with the emergence of resistance evasive strains, MIG RKN infection has the potential to become a larger issue than it already is.

As a result, there is an increasing importance in understanding the mechanisms of both MIG RKN infection, resistance, and the interaction of the *Mi* gene family with the effectome – the total set of secreted or excreted proteins intended to affect factors external to the organism - of the nematode. Not only will a greater understanding assist the creation of more resistant plant cultivars, but can also enable the development of assays to determine the potential for *Mi-1* evasion in newly detected populations.

Emergence of virulent *Meloidogyne javanica* strains

Virulence in the presence of *Mi*-hosts was initially discovered through years of repeated inoculation experiments introducing *M. javanica* VW4 onto *Mi-1* resistant tomato strains. Genetic experiments indicated that lineages VW4 and VW5 were nearly isogenic, barring one region, the *Cg-1* region which was seemingly not present in VW5, suggesting a product within the *Cg-1* region as a candidate RKN effector for *Mi-1* (Gleason, 2003). RNAi experiments to knock-down the *Cg-1* region in RKNs from the VW4 lineage enabled virulence and the ability to evade *Mi-1* resistance (Gleason et al., 2008).

Gross and Williamson (2011) described the *Cg-1* region (Figure 4.1) as 4048 bp long, flanked by the N-terminus of a gene at the 5’ (hereafter gene P) and the N-terminus of another gene at the 3’ (hereafter gene R). The region spanned between the 5' terminus of gene P and the 3' terminus of gene R is hereafter referred to as the PR region. Within the PR region a novel transposon was identified, *Tm1*, a transposable element family defined by variable-length composite terminal inverted repeats and an insertion site flanked by 9 bp repeats (Gross & Williamson, 2011).

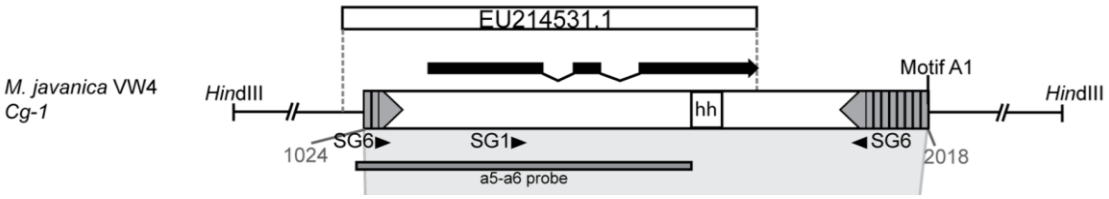


Figure 4.1: Inverted-repeat element, *Tm1*, carrying *Cg-1* and related elements (Gross and WIlliamson, 2011). Position of the *Cg-1* transcript identified by Gross and Williamson is shown with thick black lines representing exons and thin lines representing introns. Dark grey arrows within open boxes represent the TIRs with vertical bars indicating internal repeats. Target site duplications are indicated by black arrows; hh designates the position of the histone hairpin. Binding sites for primers *SG2* and *SG6* are indicated by arrowheads. The grey bar (a5–a6 probe) shows the position of the PCR amplicon used for Southern analysis.

4.1.5 Research gap, rationale, and significance of identifying candidate gain-of-virulence genes

It has been suggested that VW5 had adapted to infect resistant cultivars through transposition of *Tm1*, and that its translocation has induced functional gene loss of an effector protein detected by the host plant and *Mi-1* which in turn initiates the hosts defence pathways (Gleason, 2003; Gleason et al., 2008). This was suspected primarily because VW4 and VW5 were predicted to be isogenic with the exception of the PR region and *Tm1* (Gross & Williamson, 2011).

M. javanica has now been shown conclusively to be an allotetraploid (Blanc-Mathieu et al., 2017; Szitenberg et al., 2017; Winter et al., 2023), meaning that there are potentially four copies of the PR region in the genome. This was not known at the time of previous

studies, and has complicated our understanding of how translocation of *Tm1* could induce virulence if there are potentially three orthologous copies.

One mechanism for adaptation through transposition of *Tm1* could be gene loss. Gene loss is known to facilitate gain-of-virulence adaptations. In *Plasmodium falciparum*, loss of genes associated with antigen presentation allow the organism to evade immune detection in its host (Kirkman et al., 2014). Similar events have been detected in other parasitic species, wherein gene loss leads to loss of presentation of some factor, usually surface proteins, that then prevents detection of the infection by the host (Barry & McCulloch, 2001; Olivier et al., 2005). If *Tm1* is indeed associated with the pathway producing the protein that is recognised by *Mi-1*, loss of a copy of *Tm1* could reduce the dosage of the protein coding gene, leading to dampened recognition by *Mi-1*. In parallel, if *Tm1* is involved in the pathway but there is only one functional copy, loss of this copy would prevent or severely dampen the production of the effector protein, similarly reducing or removing the ability of *Mi-1* to recognise infection. Similarly, there may be regions in the VW5 genome that vary structurally from VW4 that could not have been detected using older short-read sequencing technologies and methods (Sedlazeck et al., 2018).

Using long sequencing reads and modern bioinformatic methods, combined with access to a phased genome assembly, we were able to compare the VW4 and VW5 genomes in the light of allopolyploidy and attain a more resolved picture. We investigated the PR region and *Tm1* to determine copy number of both the region and the gene, as well as attempting to detect any other genomic differences between the isolates. We found that the PR region is present in four copies, two in each subgenome, but that only one of these copies contains a copy of *Tm1.* Strikingly, we find a large deletion, hundreds of kilobases in length, on the copy containing the *Tm1* transposon in the virulent VW5 isolate, confirming the findings of Gleason et al. (2008) and Gross and Williamson (2011) that the *Tm1* region if absent from VW5. Alongside this, we find many more transcriptionally active genes that are deleted or disrupted in VW5 by this large deletion, generating a list of genes that are new candidates of the gain-of-virulence adaptations seen in this isolate.

4.2 Methods

4.2.1 Reproducibility

In the spirit of research integrity and reproducibility, all scripts and equivalent materials and methods will be made fully and freely available upon publication of the results of this chapter. We have adhered to FAIR principles as much as possible (Wilkinson et al., 2016). A statement of reproducibility is provided alongside this thesis. All data utilised in this study will be made available on NCBI GenBank and SRA.

4.2.2 Data acquisition and generation

In this investigation we used Iso-Seq libraries (SRA: SRX19511628) and the *Meloidogyne javanica* genome assembly (ACC: GCA_034785575.1) and corresponding annotations published by Winter et al (2024). Read libraries of VW4 and VW5 were generated in the following way. HiFi SMRTbell libraries were constructed using the SMRTbell Express Template Prep Kit v2.0 (Pacific Biosciences, Menlo Park, CA; Cat. #100-938-900) according to the manufacturer's instructions. HMW gDNA was sheared to a target DNA size distribution between 15 kb and 20 kb using Diagenode's Megaruptor 3 system (Diagenode, Belgium; Cat. B06010003). The sheared gDNA was concentrated using 0.45X of AMPure PB beads (Pacific Biosciences, Menlo Park, CA; Cat. #100-265-900) for the removal of single-strand overhangs at 37 °C for 15 minutes, followed by further enzymatic steps of DNA damage repair at 37 °C for 30 minutes, end repair and A-tailing at 20 °C for 10 minutes and 65 °C for 30 minutes, ligation of barcoded overhang adapters v3 at 20 °C for 60 minutes and 65 °C for 10 minutes to inactivate the ligase, then nuclease treated at 37 °C for 1 hour. SMRTbell libraries were purified and concentrated with 0.45X Ampure PB beads for size selection using the BluePippin/PippinHT system (Sage Science, Beverly, MA; Cat #BLF7510/HPE7510) to collect fragments greater than 7-9 kb. HiFi SMRTbell libraries were sequenced at UC Davis DNA Technologies Core (Davis, CA) using one SMRT® Cell 8M Tray (Pacific Biosciences, Menlo Park, CA; Cat #101-389-001), Sequel II sequencing chemistry 2.0, and 30-hour movies each on a PacBio Sequel II sequencer. The VW4 library was one of two generated, published, and made available on the Sequence Read Archive by Winter et al (2024; SRA: SRX19511624). All VW4 data is publicly available at the time of publication. Query sequences for genes P and R were obtained from Gleason et al (2008). Sequences for *Tm1* were obtained from Genbank (ACC: EU214531.2; Gross & Williamson, 2011).

4.2.3 Finding the PR region in the assembly

Scaffolds containing the PR region in the *M. javanica* assembly were identified using a custom *snakemake* workflow (*querying_assemblies*). This method is simple and widely used, but incorporation into a *snakemake* workflow greatly increases the ease and speed at which it can be performed. First, a *BLAST* database is generated from every scaffold in the provided assembly using *makeblastdb (-input_type 'fasta' -dbtype nucl)*. This database is then queried for a given sequence using *blastn (-outfmt 6 -max_target_seqs 50 -max_hsps 40 -evalue 1e-25)*, resulting in a list of scaffolds that contain high confidence hits to the query (Altschul et al., 1990). Sequences for gene P, gene R, and the *Tm1* transposon were provided as initial queries, followed by manual inspection of the resulting tables. Results were combined in a table (Table 4.1).

4.2.4 Describing the PR region

Coordinates of features within the PR region were identified with *BLAST* alignments of query sequences, through direct extraction from the *IGV* genome browser, or through comparison of the PR region sequences to known and published sequences on NCBI GenBank using the NCBI *BLAST* browser tool (Thorvaldsdóttir et al., 2013). A diagram model of the PR region was built using *DNAFeaturesViewer* (Zulkower & Rosser, 2020) incorporating our current findings, alongside annotations generated by Winter et al (2024) and features with high sequence similarity to published sequences. GC content was calculated from the relevant assembly scaffolds in sliding windows. Coverage was calculated through mapping of PacBio HiFi libraries to the *M. javanica* assembly.

4.2.5 Dataset generation

Read libraries of VW4 and VW5 isolates of *M. javanica* were combined using the following method to create a single fasta file of reads containing both isolates. In order to determine origin following phylogeny generation, reads were renamed according to the isolate that they originated from using a python script (*tag_read_libraries.sh*). This script preserves the identity of the read in the header, but appends the isolate name at the start. Following isolate tagging, libraries from both VW4 and VW5 were combined and the ID of each read was appended to the isolate tag for each read (*modify_entries.sh*). This produced a working dataset where each read was tagged by the isolate it originated from in order to infer tree and node content downstream.

4.2.6 Copy detection and phylogeny

A *snakemake* workflow (*copysnake*; Figure 4.2) was designed, developed, and deployed to detect homologs and homoeologs, perform alignments, and generate phylogenetic trees of query sequences from within quality controlled read libraries. This workflow generates a *BLAST* database of a read library - in this case the VW4 and VW5 PacBio HiFi libraries - removes reads shorter than a given threshold (*-min_length 4000*), and searches this database for a query sequence using *blastn* (*-max_target_seqs 500 -max_hsps 1 -evalue 0.01 -percent_identity 95*), extracting 500 reads that contain a >95% hit to the query (Altschul et al., 1990). The workflow then aligns the hit reads with *MAFFT* (*-auto -max_iterate 50*), adds and aligns the query sequence to the alignment, then trims the length of the alignment to the extremities of the query sequence (Katoh et al., 2002). Sequences containing greater than 60% gaps were removed and alignments were trimmed with *trimAL* (*-automated1;* Capella-Gutiérrez et al., 2009). The workflow then generates a phylogenetic tree with *IQTREE* (Minh et al., 2020) using the GTR model of evolution and one thousand bootstraps (*-m GTR -B 1000*), before rooting the tree at midpoint with *ete3 (Huerta-Cepas et al., 2016).* Midpoint rooted trees of the trimmed alignment are then plotted with *toytree*, *ete3*, and *ggtree* (Huerta-Cepas et al., 2016; Yu et al., 2017; Eaton, 2020). A further plot depicting the tree and its input alignment is then generated with *ggtree* and *ggmsa* (Zhou et al., 2022). By using a combined database of both virulent and avirulent isolates but retaining isolate and read information in the final alignment and tree, we can see if an isolate is absent from any node of the tree, implying absence of that locus. We initially applied this method using gene P, gene R, and *Tm1* as queries, producing alignments and trees of all copies of each within the combined database. We would later use this method on candidate genes suspected to be deleted in isolate VW5.

A variation of this workflow can be implemented that incorporates multiple queries, extracting and aligning only reads that contain both query sequences. In this way we can include intergenic regions in our alignments, enabling us to visualise the variability of these loci. Instead of trimming to the 5' and 3' of a single query, the alignment is trimmed to the start of the 5' query and the end of the 3' query. We applied this approach to the PR region, using gene P as front query and gene R as rear query, in order to extract and align reads that started at the 5' of the PR region and ended at the 3', enabling us to align and compare the intergenic region, which is predicted to contain *Tm1*.

Figure 4.2: Directed acyclic graph of the copy detection workflow, *copysnake.* Left flow shows the processing of a single query, right flow shows the processing of a bookended query, where two genes are provided and those two genes and the intergenic region are included in alignment and tree generation. Generated by *snakemake*.

4.2.7 Identification of the deleted region

Manual identification

VW4 and VW5 read libraries were aligned to the *M. javanica* assembly and manually assessed for deletions spanning the PR region using *IGV* genome browser (Robinson et al., 2011). Deletions were identified based on split mapping of VW5 reads across the PR region, partnered by a lack of split mapped VW4 reads across the same interval. Genes within this

region with corresponding mapped Iso-Seq transcripts were identified (Supplementary Table 4.1) and their sequences and headers extracted for further analysis. Notable transcripts were recorded in Table 4.2.

Read depth analysis

We applied a read depth analysis to VW4 and VW5. Reads from each isolate were mapped to the *M. javanica* genome assembly with the *vulcan* pipeline package (Fu et al., 2021). *Vulcan* is a read mapping pipeline that incorporates both *minimap2* and *NGMLR* (Sedlazeck et al., 2018). Reads are initially mapped to the assembly with *minimap2,* however in cases where the mapping score falls below a predefined threshold the read will instead be mapped with *NGMLR*. *NGMLR* handles the mapping of reads with structural variation to the reference better than *minimap2*, accounting for any structural variation between VW4 and VW5. Coverage depth scores were extracted using *samtools coverage*, which provides a count of coverage depth for each base in the reference (Li et al., 2009). For each scaffold, these counts were plotted against each other as distributions and as a sliding window of base pairs using a script obtained from Winter et al (2024; *depth_analysis_setup.R, depth_analysis.R*) to detect regions where relative coverage differed, implying a difference in copy number or indel distribution between VW4 and VW5 (Figure 4.7).

4.2.8 Ontological annotation and copy analysis of deleted genes

Sequences and annotations corresponding to transcripts that mapped between the coordinates of the identified deletion were extracted from the Iso-Seq annotations of Winter et al (2024) and collated. Extracted transcripts were input as queries to the NCBI RefSeq protein database using *BLAST* (Altschul et al., 1990) in order to identify similarity to published proteins and to assign gene ontology and functional annotation, results of which were assigned for each (Supplementary Table 4.1). In order to identify potential paralogs of deleted or interrupted loci, extracted transcripts were entered as queries into the *copysnake* workflow, which identified all copies of given query genes in our read libraries and generated phylogenetic trees representing them. Alongside this, query genes were compared against the *M. javanica* assembly using the *querying_assemblies* workflow to identify the amount any locations with high nucleotide similarity. Through combination of these two results we can infer if there are paralogs of the genes deleted on scaffold 15 elsewhere in the genome.

4.3 Results

4.3.1 Finding the PR region in the *M. javanica* assembly

Through local alignment of query sequences we identified two scaffolds in the *M. javanica* genome assembly containing genes P and R, scaffolds 15 and 13, which form a homoeologous chromosome pair (Winter et al., 2023). We found no other copies of genes P or R elsewhere in the genome. This suggests that there are at least two copies of the PR region within the genome, one on each subgenome, with both P and R genes around 4 Kbp apart. We find one copy of *Tm1* on Scaffold 15 positioned in the intergenic region between P and R, and no copies on Scaffold 13. However, we do identify copies of *Tm1* on Scaffold 18 and Scaffold 11, with 98.6% and 97.6% similarity to the query sequence, and 98% and 99% query cover, respectively (Table 4.1).

Table 4.1: Best *BLAST* hits of query genes P and R against the *M. javanica* assembly.

Query	Hit	Identity (%)	Query cover (%)	Start coordinate	Stop coordinate
Gene P	Scaffold 15	100	100	2,896,333	2,893,921
Gene P	Scaffold 13	91.142	35	2,810,288	2,809,461
Gene R	Scaffold 15	100	100	2,889,583	2,887,901
Gene R	Scaffold 13	90.049	70	2,806,001	2,804,785
Tm1	Scaffold 15	100	100	2,891,298	2,890,304
Tm1	Scaffold 18	98.6	98	1,181,282	1,180,319
Tm1	Scaffold 11	97.6	99	1,858,627	1,859,611

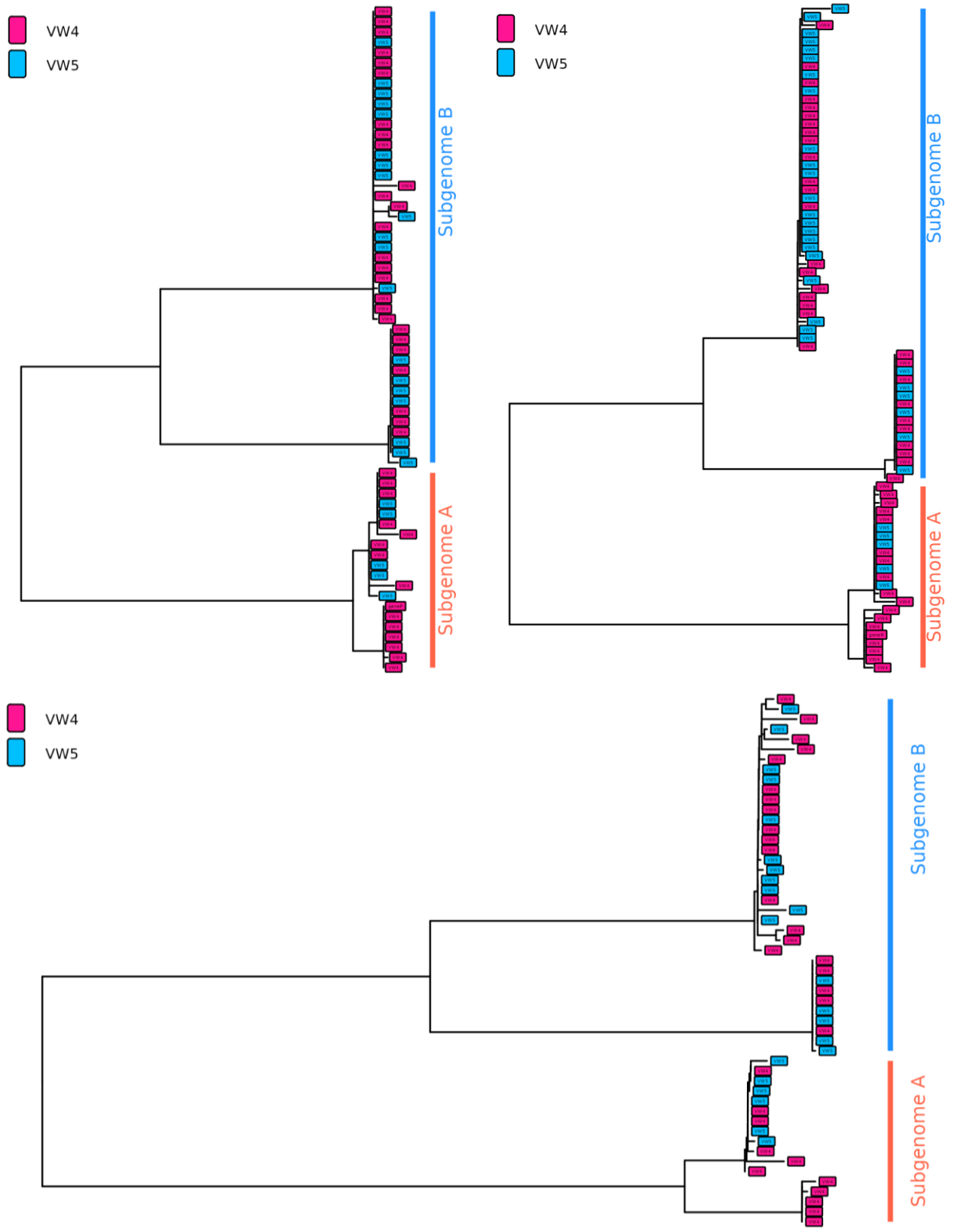


Figure 4.3: Phylogenies generated from copies of genes P and R in the trimmed read libraries. Four clear clusters are present for both, with similar branch length and topology. Reads from VW5 are completely absent from one cluster, indicating that this copy is deleted in VW5. Left, Gene P. Right, Gene R. Bottom, PR region.

These findings indicate that the PR region is tetraploid in *M. javanica*, and that the four copies share an evolutionary history suggesting the hybridisation of two diploids to form the allotetraploid species. It also indicates that isolates VW4 and VW5 are not isogenic across this region, with a deletion spanning, at the least, from the 5' terminus of gene P to the 3' terminus of gene R. The full extent of this deletion was investigated in greater depth.

4.3.2 Describing the PR region

Visualisations in *IGV* of the PR region on these scaffolds as well as visualised alignments generated by the *copysnake* workflow show four distinct copies within VW4, two copies for each scaffold. The *Tm1* transposon is present within the PR intergenic region of one of the copies mapped to scaffold 15, hereafter copy A1. The intergenic regions of the other copies are much shorter than the *Tm1*-containing intergenic region of A1, and all exhibit unique sequences differing from that of other copies.

Comparison of gene P and gene R against the RefSeq Protein database suggests that they are a phosphatidylinositol transfer protein (PITP) and a Rho GTP-ase activating protein (RhoGAP) respectively. We find a transcript with high sequence similarity to gene R in the transcript annotations of Winter et al (2024) that corresponds to Rho-GTPase activating domain in the RefSeq database. Rho GTPases are involved in cytoskeleton maintenance and formation (Etienne-Manneville & Hall, 2002).

We do not find a transcript correlating to gene P in these annotations. We do however find a correlation to a gene in the gene predictions of Winter et al. (2024), suggesting that gene P does not produce a transcribed product, or that this product is not present in our IsoSeq library. When assigned ontology with *BLAST* this sequence is identified as a possible phosphatidylinositol transfer protein. Phosphatidylinositol transferase functions as a mode of transport for phosphatidylinositols, which play important roles in cell signalling and membrane trafficking.

A model of the PR region was built using annotations generated by Winter et al (2024; Figure 4.5). We found that there are four distinct alleles of the PR region, two assigned to each subgenome of *M. javanica*. All copies have very low nucleotide divergence in the coding regions of genes P and R. Copies show distinction in the intergenic region, with each copy containing a distinct sequence. The *Tm1* sequence can be found in one allele of the PR region on subgenome A in VW4.

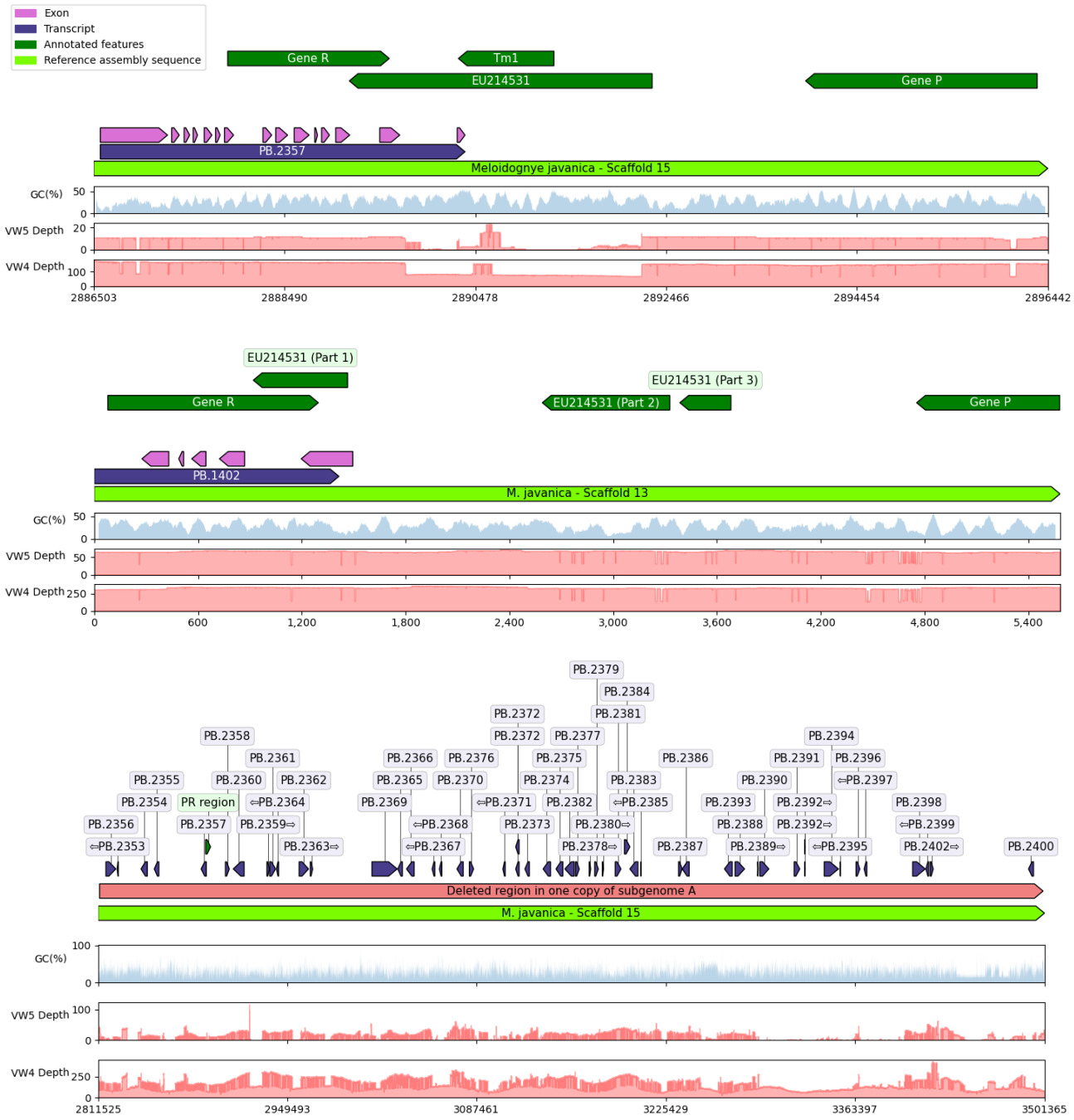


Figure 4.5: Diagrams of the PR region in scaffold 15 (top) and scaffold 13 (middle) of the *M. javanica* genome assembly, and the deleted region of subgenome A identified on scaffold 15 (bottom). Dark green features are regions of sequence similarity with sequences identified by Gleason et al. (2008) and Gross and Williamson (2011). Red features indicate deletion. Lime indicates the reference sequence. Purple indicates Iso-Seq transcripts generated by Winter et al. (2024). Pink indicates exons. Blue histogram is GC content. Red histograms are coverage of the scaffold within VW4 and VW5 read libraries. Generated using *DnaFeaturesViewer.* A region of decreased coverage can be seen within

the PR region of scaffold 15 of VW4, indicating that this region is only present in one copy. In VW5, coverage of this region falls to near zero, indicating that the single copy found in VW4 is not present in VW5. The bottom image shows Iso-Seq transcripts located within the deleted region, as well as the relevant location of the PR region within it. There are several regions across this sequence where coverage of VW4 halves, suggesting deletions between each copy of the subgenome. In VW5, these regions fall to zero coverage, supporting that some sections and their corresponding transcripts are completely absent from VW5. These genes are recorded in Supplementary Table 4.1. Genes appearing to be completely deleted in VW5 are recorded in Table 4.2.

4.3.3 Identification of deleted regions

Manual identification

Following discovery that the PR region of the A1 copy was deleted in VW5, we investigated the extent of this deletion. We identified a larger deletion exclusive to the A1 copy of VW5, spanning ~687 Kbp and encompassing the PR region as well as several other genic and coding regions (Supplementary Table 4.1), which we extracted from the assembly as potential candidates for gain-of-virulence adaptations through gene loss. Some of the genes spanned by the deletion also lack an A2 homolog, indicating that these loci have been fully deleted from this homoeolog. There are also several genes that although not fully deleted, are disrupted by deletions, with at least one exon included in the deletion.

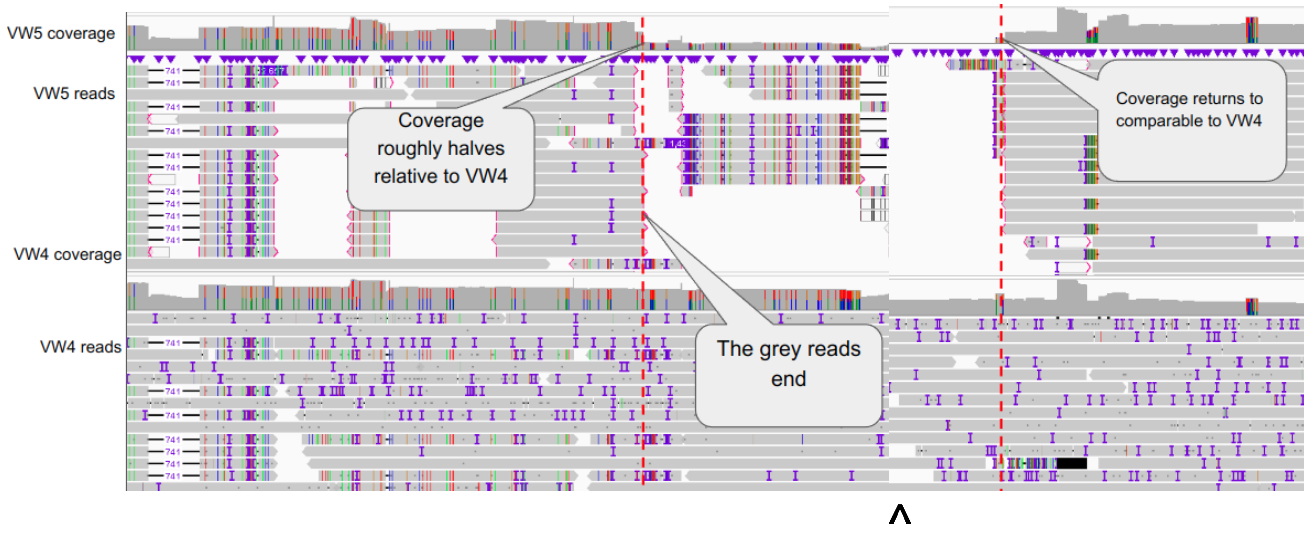


Figure 4.6: IGV screenshot showing the start and end sites of the deletion. Cropped, internal view removed for visualisation (^). Produced in *IGV*. All sequences mapped against *M. javanica* reference assembly (Winter et al., 2024). Histogram at top of each read track

indicates coverage. Grey bars in reads tracks indicate reads. Colours within grey reads track indicate divergence of mapped read from reference. In the VW5 reads track, no reads can be seen spanning the region within the boundaries of the suggested deletion, in contrast to VW4 where reads map continuously throughout. Coverage in VW5 falls to almost nothing within suggested deletion, whereas in VW4 it remains consistent throughout. We observe reads sequenced from VW5 spanning the area of the presumed deletion, split mapping strongly to either side.

Read depth analysis

When visualising read mapping depth of both VW4 and VW5 libraries, it is clear that there is a large, proportional drop in coverage depth between the same coordinates we identified through visualisation of split-read mapping (Figure 4.7). No similar drop was seen for any of the other scaffolds in the assembly. This deletion starts around 75 Mbp upstream and around 580 Kbp downstream of the PR region, totalling a deletion of ~687 Kbp in one copy of scaffold 15, at coordinates 2,812,526 to 3,500,365. This corresponds to the split mapped reads seen previously (Figure 4.6).

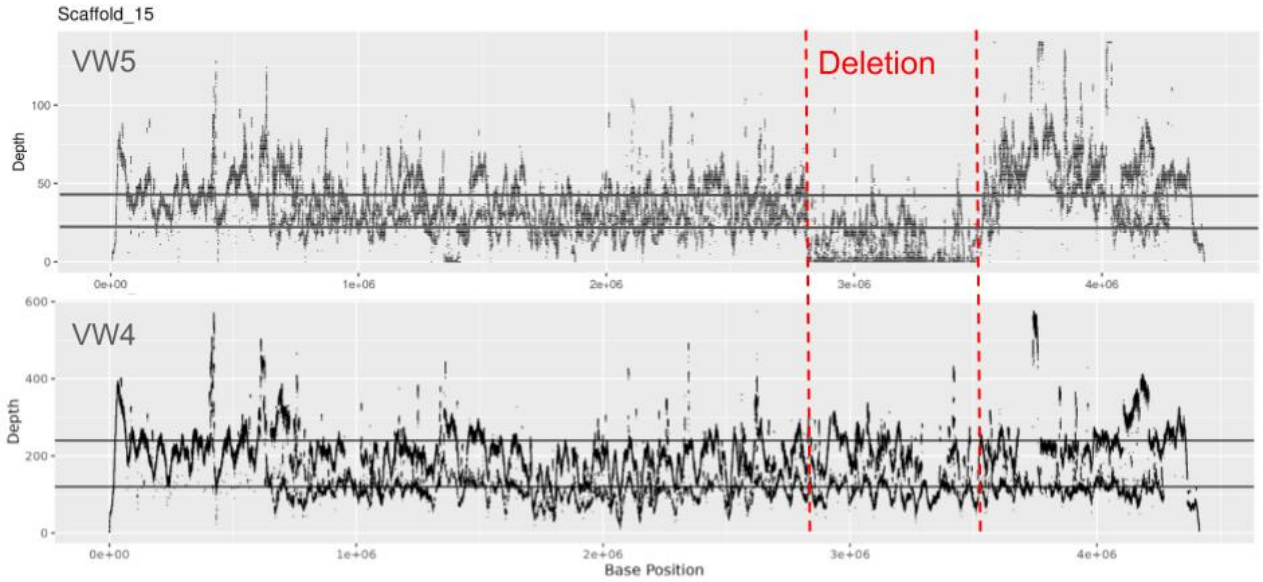


Figure 4.7: Differences in coverage depth on scaffold 15 overlaps with the region delimited by split reads exclusive to VW5. Plot shows coverage (in 1000 bp windows) of mapped reads along the chromosome identified to contain the PR region and the Tm1 transposon. The region spanned by split reads identified within IGV is marked in red, which coincides with a reduction of coverage depth to near zero in VW5.

The majority of this region is tetraploid in VW4 and triploid in VW5, however a section of the deletion spans a region that is triploid in VW4, with only one A copy. In VW5 this region is diploid, with no A copies at all, suggesting that several genes on this homoeolog could be entirely deleted. Information about transcripts mapped within this region were added to the table of candidates for phenotype changes between VW4 and VW5 (Supplementary Table 4.1) and genes identified as deleted or disrupted were recorded in Table 4.2.

4.3.4 Ontological annotation and copy analysis of deleted genes

For each gene identified as missing in the deletion, we applied our *copysnake* workflow. Selected trees are shown in Figure 4.8. All trees can be found in Supplementary Figure 4.1. Table 4.2 describes the genes within the deleted region with notable functional annotation and the corresponding best *BLAST* hit. Of note is transcript PB.2397, which was identified through functional annotation to be papain-like cysteine protease (PLCP) cathepsin B. This gene has zero coverage in VW5, and phylogenetic analysis indicates only 3 copies within VW4 and two in VW5, both of which are on subgenome B. PLCPs are heavily involved in both plant host defence response, as well as infection capability in RKNs. A similar and nearby gene, PB.2396, is also identified as a cathepsin B like isoform, however this gene is not fully deleted in VW5. Of the four exons in this gene one is at zero coverage, suggesting that this gene is not wholly deleted, but disrupted in a way that would prevent functionality of the protein product. Average coverage depth of VW5 mapped against the Winter et al (2024) assembly is 45.83x, compared to an average coverage depth of 216.33x in VW4, a ratio of around 4.7:1. This is important to consider when comparing coverage depth of loci within the deletion, as any loci that deviate from this distribution are likely disrupted or deleted in VW5.

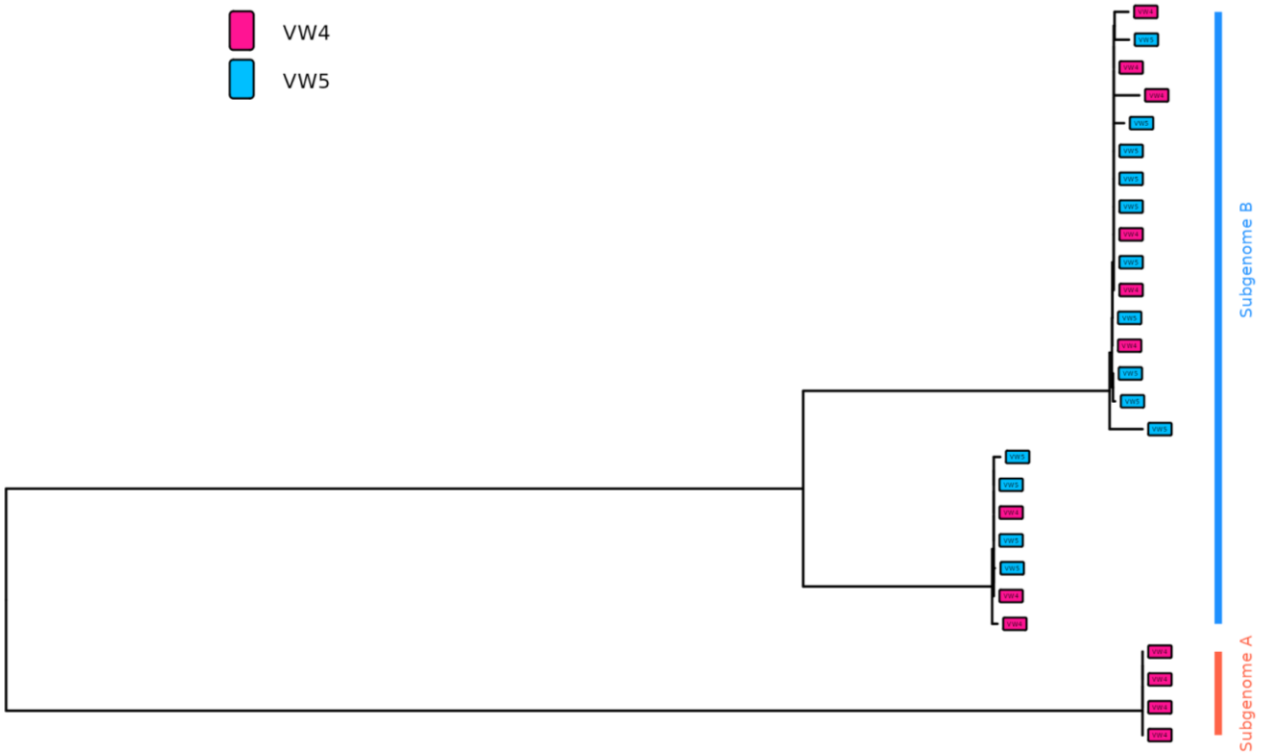


Figure 4.8: Phylogenetic trees of all copies of PB.2397, a suspected PLCP, found in the combined libraries of VW4 and VW5. Three clusters are seen for VW4, indicating that there are only three copies in VW4. No subgenome A copy of this gene is detected in VW5.

Table 4.2: Notable genes identified as deleted within the confines of the identified deletion.

Transcript	Status	Best BLAST hit	Coverage ratio (VW4:VW5)	Notes
PB.2367	Deleted from subgenome A	Hypothetical protein GCK72_008905	138.29:1	N/A
PB.2368	Deleted from subgenome A	Hypothetical protein GCK72_008905	N/A	Zero coverage in VW5
PB.2390	Deleted from subgenome A	Integrative alpha pat-2 precursor	22.11:1	Zero coverage
PB.2391	Likely deleted in subgenome A	Hypothetical protein LOAG_16918	77.56:1	Half of exons at 1 coverage, rest at 0
PB.2392	Deleted from subgenome A	N/A	N/A	Zero coverage in VW5
PB.2394	Disrupted	Hypothetical protein LOAG_16918	20.40:1	Several exons at zero coverage

PB.2395	Likely deleted in subgenome A	N/A	108.39:1	Coverage is <1
PB.2396	Disrupted	Cathepsin B	10.69:1	One of four exons at zero coverage
PB.2397	Deleted in subgenome A	Cathepsin B-like isoform X1	N/A	Zero coverage in VW5
PB.2400	Disrupted	Aldehyde dehydrogenase domain-containing protein	7.05:1	One exon at zero coverage

4.4 Discussion

We have shown that the PR region and *Tm1* are indeed missing in VW5. Alongside, confirming some results of Gleason (2008) and Gross and Williamson (2011), we have revealed that the reason for this absence is that a much larger deletion is present in VW5, spanning over ~680 Kbp of one allele of subgenome A and encompassing the PR region. We characterised this important region for the first time, describing how there are 4 copies of the PR region in VW4, with *Tm1* present in the intergenic region of only one of these copies, and that VW5 contains only 3 copies of the PR region, with the copy containing *Tm1* having been deleted. We also identified up to 50 transcriptionally active genes that are impacted by this deletion, either being reduced in copy number, disrupted by loss of one or more exons or entirely deleted and not present at all in the VW5 lineage of *M. javanica*. We functionally annotated these genes and will attempt to provide context into how their disruption could enable the VW5 lineage of *Meloidogyne javanica* to evade *Mi-1* resistance in their hosts.

An alternative hypothesis to that of a large deletion could be the presence of highly divergent haplotypes or regions within *M. javanica*, like those observed in other nematode species, notably *C. elegans, Heligmosomoides polygyrus* and *H. bakeri* (Lee et al., 2021; Stevens et al., 2023). In this proposed system, the absence of coverage of one allele of VW5 could be explained by high divergence between the VW5 haplotype and the haplotype of the VW4 *M. javanica* reference assembly causing sequenced reads to fail to map, leading to the apparent deletion observed here. Although we cannot rule out that *M. javanica* or other MIG species may contain genomic regions exhibiting highly divergent alleles - and we anecdotally observed what appeared to be highly divergent intergenic regions in some cases - we believe that this hypothesis can be rejected as an alternative explanation for the

observed deleted region in VW5. First, the region is ~650 Kbp long. If the lack of reads mapping was due to high divergence between isolates, this divergence would have to arise incredibly fast, as VW4 and VW5 did not diverge long enough ago to accumulate such a large divergence. Second, we observe PacBio HiFi reads from our VW5 genomic library split mapping to either side of the deletion, indicating strongly that the proposed deleted region is indeed missing in that isolate, and is not in fact showing no coverage due to high divergence. We therefore refute this alternative hypothesis as an explanation for the apparent deletion.

4.4.1 Advances in understanding the PR region, Tm1, and differences between VW4 and VW5

We have successfully and thoroughly investigated and annotated the PR region in *Meloidogyne javanica*, finding the region present in four copies in VW4; two to each subgenome, on scaffolds 15 and 13. We also find the *Tm1* transposon present on the A1 copy of scaffold 15. Gleason (2008) found that VW4 and VW5 lineages were nearly isogenic with only one difference in the PR region; a suggested deletion caused by transposition of the *Tm1* transposon. Though we have confirmed that the *Tm1* transposon is not present in VW5, we find that this is not due to movement of the transposon, but due to a large ~680 Kbp deletion on the A1 copy, spanning the PR region. This is the only large-scale inconsistency we could detect between the VW4 and VW5 genomes. We do not find that transposition of *Tm1* interferes structurally with any coding transcripts, nor does it contain any transcriptionally active coding sequences.

4.4.2 Gene loss could prevent host recognition of infection

Mi-1 recognises a secretory protein (effector) released by the RKN, initiating a defence cascade in the host (Milligan et al., 1998; Gleason et al., 2008). Loss of the effector would prevent activation of *Mi-1*, preventing the host from initiating defence mechanisms, allowing VW5 to infect and reproduce on hosts normally resistant to *M. javanica*. We find candidate genes for effectors of *Mi-1* within the VW5 deletion, including two transcripts that were identified as cathepsin-B papain-like cysteine proteases (PLCPs). A sensible continuation and validation of this work would first be to attempt to detect this gene, as well as other genes identified in this work as being deleted or disrupted in VW5, using PCR amplification. Early investigation in this direction has been attempted, and shown that this

gene is unable to be detected in VW5 using this approach (Siddique, pers comms.), indicating the validity of the findings presented here, and strengthening the support for this gene being fully deleted in VW5. Further than this, knock-down or knock-out experiments could be designed to interfere with this gene in isolates of VW4 and determine if doing so leads to evasion of *Mi-1* engineered resistance or origination of gain-of-virulence phenotypes.

PLCPs are thought to be central to plant immunity and immune response coordination, and a mounting body of evidence is indicating that the targeting of them by parasite effector proteins is a frequent and important mechanism by which the parasite can increase the chance of establishing an infection and suppress host resistance and immune responses (Shindo & Van der Hoorn, 2008; Misas-Villamil et al., 2016). Alongside being important in host defence pathways, PLCPs are secreted by the nematode and are involved in tissue penetration and feeding site formation (Shindo, 2010; Mejias et al., 2019; Pogorelko et al., 2019). As the VW5 isolate can still infect the host even without these genes, it does not appear that they are integral to the infection process, although it has been observed that VW5 has reduced virulence on non-resistance host cultivars when compared to VW4 (Siddique, pers comms.). This is suggestive of a coevolutionary conflict, where a mutation that confers reduced fitness in some cases con confer a large fitness gain in a niche case, in this case the infection of resistant cultivars (Kraaijeveld et al., 1998; Jones & Dangl, 2006; Betts et al., 2016). It may be that a protein within the host - or *Mi-1* itself, given that this mutation enables evasion of *Mi-1* - relies on presence and subsequent detection of this transcript or translated protein to initiate the defence cascade. In the absence of this initiator the plant is either unable to detect infection or unable to deploy resistance mechanisms.

Loss of effector genes in plant pathogens, leading to adaptation and subsequent resistance evasion, is well documented in bacteria (Reddick & Alto, 2014; Hemara et al., 2022) and fungi (Dodds et al., 2009; Fouché et al., 2018). Based on the findings presented here, we propose that evasion of *Mi-1* induced resistance in VW5 likely arose from a similar mechanism, with the deletion of the effector - or an essential gene in the effector pathway - preventing activation of the *Mi-1* resistance pathway in the host, enabling infection. The exact mechanism through which *Mi-1* mediated resistance is activated is unknown (Pascual et al., 2024), however with the only known genomic differences between VW4 and VW5 being the deletion identified here, it follows that loss of one or more of the genes in this region is preventing secretion of the *Mi-1* associated effector.

4.4.3 Identification of definitive mutation enabling *Mi-1* evasion can enable fast assessment of virulence

The current method for determining virulence or avirulence of the newly acquired population of *M. javanica* is through the use of laboratory growth experiments, where the population is isolated and allowed to colonise a resistant host, the success or failure of which denotes the virulence (Castagnone-Sereno, 1994; Williamson, 1998). Several studies have now been performed with RKN wherein population demography can be determined through the use of quickly available PCR assays (Sellers et al., 2021). The identification of a deletion definitively associated with gain-of-virulence mutations in *M. javanica*, specifically one bestowing the ability to evade *Mi-1* engineered resistance, can lead to the development of PCR assays to assess virulence, reducing the diagnostic time from weeks to a few hours.

4.4.4 Detecting deletions and deleted candidates in other MIG species

Further studies could apply the same approach used here to other MIG species with known and established virulent lineages. Aside from being a tested method of detecting differences between lineages, identification of similar genes absent from virulent lineages but present in avirulent ones could add weight to the candidates suggested here, as the effector mechanism of *Mi-1* mediated resistance is expected to be applicable to all MIG species. Applying this approach to other MIG species could validate candidate genes in *Mi-1* resistance evasion.

4.4.5 Future validation and comparison

Future validation of deletion of the genes identified here with PCR will further support our findings. Genes suspected to be deleted in VW5 should be amplifiable in VW4 but not VW5. If the primary candidates are confirmed to be missing in VW5 with PCR, further tests could include the knockout of those genes in VW4 to determine if their loss does indeed affect the virulence phenotype.

The *copysnake* workflow could be applied to other species of *Meloidogyne* with lineages able to evade *Mi-1* mediated resistance using some of our detected candidates as

queries. If copy number differences are also identified, it could identify a convergent cause of resistance evasion, as well as identifying candidates for gain-of-virulence in those species.

Identification of the function and structure of the resulting protein products of transcripts identified here and confirmed to be missing through PCR could further elucidate what possible pathways or interactions they are involved with.

Since the research and results presented in this chapter were completed, two assemblies of *M. javanica* were published (Dai et al., 2023; Mota et al., 2024). Although reanalysis of this extra data has not been possible here, opportunities for additional analyses are included in the General Discussion chapter (Chapter 5).

Chapter 5: General Discussion

Michael R Winter[1]

[1] School of Natural Sciences, University of Hull, Kingston-upon-Hull, UK.

Correspondence: mrmrwinter@gmail.com

5.1 Overall summary of thesis

We set out to assemble the genome of *Meloidogyne javanica* to a higher standard of quality than had previously been achieved, phase this assembly into its constituent subgenomes, and perform a comparative analysis between those phased subgenomes. We then hypothesised that due to the orthologous nature of the subgenomes of other MIG species, it would be possible to phase assemblies of other species using our novel assembly, and perform comparative genomic analysis across the group. Among the available long read clade I *Meloidogyne* assemblies, *M. luci* appeared to be a good candidate for inclusion in this analysis, and as such was included in further analysis. Alongside performing comparative genomics of the group, we also sought to use our assembly to investigate the suspected causes of gain-of-virulence mutations in the VW5 lineage of *M. javanica*, investigating the region thought to be divergent between isolates.

In all areas of this work we tried to use the most robust and modern technical methods available to us. This included sequencing technologies such as PacBio HiFi and Iso-Seq, ultra-long read Oxford Nanopore sequencing, or Hi-C chromatin contact conformation data, as well as highly regarded and robust bioinformatic methods such as *MAKER* and *HyPhy*, to phasing with repetitive ancestral k-mers. Alongside this we develop several novel methods and bioinformatic workflows, all of which are made freely available for other researchers.

Other work

The analyses contained in Chapters 3 and 4 of this thesis were completed by December 2023, and Chapter 2 was written as a manuscript and submitted to *bioRxiv* and *PLOS One* in May 2023. In the 13 months before publication of this manuscript by *PLOS One* in June 2024, two assemblies of *M. javanica* were published by other research groups (Dai et al., 2023; Mota et al., 2024). The assembly published by Mota et al. (2024), although valuable in many ways, was not chromosomal, assembling *M. javanica* into several hundred contigs and not phasing those contigs into subgenomes. The work by Dai et al. (2023), in contrast, claims to have assembled the *M. javanica* genome to chromosome-scale and fully phased the assembly into each subgenome copy (A1, A2, B1, and B2) as well as to have carried out molecular evolution and some inter-species comparative analysis as we have in this thesis. This work took a consortium approach with large amounts of funding and expertise applied, and as such their assemblies appear to be of very high quality. Due to the time disparity between consolidation of our findings and this publication, it has not been

addressed in the data chapters above, however its relevance to the discussion of our overall findings is included below.

Two papers were published at a similar time to the submission of Chapter 2 to PLoS One, which describe different genome assemblies of *M. javanica*. A comparison of these assemblies to the assembly presented here in Chapter 2 was requested, including a pairwise alignment and statement of collinearity to determine the congruency of the results. Using *DGENIES*, a pairwise alignment and visualisation software, we aligned and compared our *M. javanica* assembly to those of Mota et al (2024) and Dai et al (2023). We find the following. For both Mota et al and Dai et al assemblies, we observe an alignment rate of 53.22% and 51.64% respectively, combined with strong collinearity of aligned sequences. This is congruent with our assembly being a haploid representation of a tetraploid (2n = 2x) and theirs being a haplotype resolved representation (2n = 4x). Of the remaining aligned percent, our assembly aligns at >75% similarity with 36.16% of the Mota et al assembly (>75% = 36.16%, <75% = 14.72%, <50% = 1.89%, <25% = 0.44%), and 38.45% of the Dai et al assembly (>75% = 38.45%, <75% = 11.51%, <50% = 1.23%, <25% = 0.45%).

5.2 Contributions to the field and interpretations

5.2.1 Assembly

We found that through the application of modern approaches it was possible to generate a chromosome-scale assembly of an allopolyploid MIG species as well as phase and assign the majority of it to either subgenome, which was a first for allopolyploid species in the *Meloidogyne incognita* group (MIG). Genes and repeats in the assembly were annotated using high quality transcriptomic data and modern bioinformatic methods.

For future assemblies of allopolyploid genomes, we would potentially use a different strategy. When the initial draft assemblies were generated, we aimed for contiguity and preservation of subgenomes over overall copy content, ie, a dual assembly. That is, a pseudo-diploid assembly (Li, 2021), where each copy is a haploid representation of a subgenome (A/B; homoeologs). One reason for this is that we were unsure how successful scaffolding using Hi-C would be in the presence of four copies (homoeologs and haplotigs, ie, all haplotypes). In the time since the assembly has been finished, the use of chromatin conformation capture sequencing to scaffold allopolyploids has become more frequent and

resources guiding the performance of such a method more widespread, and it has become clear that it is possible to use it both to scaffold an assembly and to identify and separate homeologous contigs and their individual copies. With this in mind, during future allopolyploid assemblies I would aim to retain as many contigs from the draft assembly - representing all haplotypes - as possible, avoiding the purging of duplicates and haplotigs, then go on to scaffold with ONT and Hi-C to attain chromosome-scale. Phasing into subgenomes can then be performed, generating a haplotype-resolved assembly (Li, 2021).

An alternative method of scaffolding haplotype-resolved allopolyploid assemblies could be performed through the use of ultra-long Oxford Nanopore Technology reads. In theory, these reads should adequately span shorter unscaffolded contigs, with enough sequence divergence between haplotypes to prevent phase switch errors being introduced, however it must be said that this is highly dependent on both the quality of the ultra-long reads, the absence or previously existing phase switch errors in the contigs to be scaffolded, and a high enough amount of coverage of scaffold points to ensure accuracy and reliability. If parameters for alignment and scaffolding were correctly optimised, it should be possible to generate chromosome scale scaffolds without requiring chromatin contact information.

5.2.2 Phasing other assemblies

We then used this newly produced *M. javanica* assembly to successfully phase the assemblies and annotations of other allopolyploid *Meloidogyne* species, including *Meloidogyne arenaria, Meloidogyne incognita,* and *Meloidogyne luci*. These outputs are now publicly available on NCBI Genbank and Zenodo (10.5281/zenodo.13819447), providing up to date and high quality genomic resources for further investigation of this species and this group in future generations of genomics. Phasing using repetitive ancestral k-mers, as described in Chapter 2, depends on all scaffolds containing some of these k-mers, the probability of which is inversely correlated to levels of fragmentation of the target assembly. In the future we expect that chromosome scale assemblies of other MIG species will be produced, the contiguity of which will enable repetitive k-mer based phasing.

Other methods of phasing subgenomes include using synteny, as performed by Dai et al. (2023), whereby scaffolds of allopolyploid species are assigned to either A or B subgenomes through shared synteny with a diploid relative, in this case *Meloidogyne graminicola.* Synteny based methods of phasing genome assemblies rely on closely related evolutionary histories of species used as reference, as well as a shared common ancestor

that is not too distant. It also assumes a negligible amount of homoeologous gene conversion. Our opinion is that the ancestral k-mer based phasing method is superior, however it is dependent on contiguous assembly, whereas the synteny approach is potentially more tolerant of fragmentation.

5.2.3 Multispecies comparison

We have performed the first multispecies comparison of the subgenomes of the MIG and *Meloidogyne luci.* We establish the ploidy state of these species through application of k-mer based techniques, finding that *M. luci* and *M. incognita* are triploid and that *M. javanica* and *M. arenaria* are tetraploid, and provide these results alongside up to date metrics of their genome assemblies and recent genomic resources. These metrics include descriptive statistics such as length, scaffold number, N50, and production method, as well as estimates of k-mer uniqueness and duplication rate. Similar collections of descriptive data have been published in the past (Szitenberg et al., 2017; Koutsovoulos et al., 2020), however ours contains the most up to date information at the time of analysis.

5.2.4 Phylogeny

We find that the phylogenetic position of the two subgenomes of *M. luci* are paraphyletic, with each subgenome more closely related to its homeologous copies in other species than one another, strongly suggesting a hybrid origin of *M. luci* and a shared parentage with the MIG. This raises questions about the evolutionary history of *M. luci* and its subgenomes. If *M. luci* originated from the same hybridisation event as the MIG, then *M luci* diverged from the common ancestor of the MIG before that MIG ancestor diverged. One explanation for this is a greater skew of subgenome dominance being present between the subgenomes of *M. luci*, wherein one subgenome was able to establish dominance over the other, leading to loss of a non-dominant copy and reversion to triploidy (Wolfe, 2001; Mandáková et al., 2016; Bird, 2022). Chromosome loss occurring on one subgenomic copy, leading to a hypotriploid or hypotetraploid state, has been predicted in allopolyploid *Meloidogyne* species (Triantaphyllou, 1981, 1985; Abad et al., 2008; Szitenberg et al., 2017; Despot-Slade et al., 2022)

A similar type of event could be what delineated *M. incognita* and *M. floridensis* grouping from the *M. javanica* and *M. arenaria* grouping. There is some evidence for such a situation in our results. For example, *M. luci* displays a much lower percentage of unique k-

mers than do the MIG species, suggesting a lower amount of divergence between copies. Alongside this is the difference in relative rate of mutation between subgenomes of *M. luci* when compared to the MIG species. All MIG species tested displayed a significant difference in substitution rates between transcriptional regions of subgenomes, however a significant difference was not seen between subgenomes of *M. luci.* These things together could be indicative that *M. luci* subgenomes have resolved any subgenomic conflict and have "settled", i.e., the initial genomic instability and subsequent process of gene fractionation has been resolved. We would hypothesise that the ancestor of *M. luci* experienced a faster onset of subgenome dominance than the ancestor of the MIG species, leading to loss of a subgenomic copy, an increased rate of homeologous conversion and subsequent homogenisation. Literature investigating the temporal interaction of heterozygosity and subgenome dominance is minimal, though it would be expected that as time progresses continued homoeologous conversion combined with non-functionalisation of non-dominant loci will work to reduce overall heterozygosity of orthologous transcriptional regions.

This analysis can be expanded upon through the inclusion of other lineages of the species included here. For example, the isolates used in the Dai et al (2023) and Mota (2024) assemblies of *M. javanica, M. arenaria,* and *M. incognita* are distinct from those included in our analysis in Chapter 3. Inclusion of these isolates will increase the robustness of a genome scale phylogeny. Furthermore, as more genomes of clade I species of *Meloidogyne* are assembled to a chromosome-scale, they too can be included in such an analysis. We hypothesise that all clade I species derivative of *M. luci* will contain orthologous subgenomes and the same parent species.

5.2.5 Synteny

Highly fragmented, unphased assemblies have made meaningful detection of synteny between subgenomes difficult (Bhutkar et al., 2006; Liu et al., 2018), but through applying our chromosome-scale assembly of *M. javanica* as an anchor, we have been able to detect large amounts of synteny between homeologous subgenomes across species. If this is the case, then it would be possible and somewhat reliable as a model to pseudo-scaffold fragmented assemblies of closely related MIG species using the *M. javanica* assembly as reference. This would enable a deeper investigation of structural differences between these species. As well as this, it would be interesting to see how far through the clade I *Meloidogyne* species this trend continues. Alongside this we observe synteny and collinearity between subgenome A and subgenome B in *M. javanica* in substantial blocks

with few rearrangements. This suggests that the parental species of these hybrids had maintained the same if not greater levels of synteny and collinearity prior to the hybridisation event but after their speciation. The parental species being closely related in both sequence and karyotype was probably a contributing factor to the success of the hybridisation. Investigations of synteny are important to understanding the origins and evolution of a group of species or genomes (Tang et al., 2008; Ijaz et al., 2022). It also allows us to see large structural differences as well as potential chromosome fission or fusion, things that are expected to play roles in the genome evolution of allopolyploid species. This is of particular interest in the MIG, where striking variation in chromosome number and karyotype between species and subgenomes have been observed (Eisenback & Triantaphyllou, 1991; Carlton et al., 2022; Dai et al., 2023; Blaxter et al., 2024).

5.2.6 Subgenome dominance

With phased assemblies of several allopolyploid *Meloidogyne* species, and in the knowledge that subgenomes of these species shared parental origins, we were able to perform tests for signals of subgenome dominance in these species for the first time. We find that although there are signs of subgenome A being more dominant than subgenome B, these signs are slight, and given the persistence of polyploidy in these species, subgenome dominance is likely balanced enough that both subgenomes are being retained. An option for definitively testing for an inverse relationship between omega values of subgenome A and B in a given species, and potentially selection pressure, would be to test the distributions presented in Figure 3.10 for significant differences. The null hypothesis for this test would be that there is no difference in distribution of omega between A and B subgenomes in an orthogroup. The alternative hypothesis would be that the distribution of omega values for A and B subgenomes is inversely correlated, and that a lower omega in one is associated with a raised omega in its homoeolog. The abnormal distribution of omega values excludes parametric tests, leaving the following non-parametric options. Both Mann-Whitney *U* and Kolmogorov-Smirnov tests would enable us to determine if the distributions differed significantly from one another (Hettmansperger & McKean, 2010; Berger & Zhou, 2014). If a significant difference is determined, the inverse correlation could be tested using the Spearman's rank correlation coefficient (Gauthier, 2001). Linkage of genes could introduce skew to this analysis, but can be managed through exclusion of genes/orthogroups that are located close to each other along a chromosome. The most contiguous assembly in the analysis should be used for this, as it has been shown that these

species retain synteny and collinearity, so it would be safe to assume that locus coordinates are consistent across species. Alternatively, a bootstrapping approach could be taken with a random subset of orthogroups withheld from each iteration in order to achieve an aggregated result.

Future research can repeat our methods for testing mutation rates - both subgenome to subgenome and at individual loci - but utilising different models that may be more sensitive and accurate in the detection of low omega and potential purifying selection. Although valid for identifying differences between omega at a given locus, the model we use in Chapter 3 is designed primarily for detection of positive selection, so utilising a model where neutral selection is the null hypothesis would perhaps give more accurate results in terms of loci with low omega. Model selection can be complicated, and we feel that none of the models available were a great fit considering the evolutionary history of the taxa analysed, however there is likely no way around this as models are developed primarily around model organisms with simpler reproductive strategies.

Although we do not find a strong enough swing of dominance toward either subgenome based on mutation rate to declare one dominant over the other, we do find incidences at a gene level where one subgenome appears to be dominant. Further investigation into subgenome dominance and its effect on genome evolution in clade I allopolyploids could take several roots. Dominance is believed to be linked to specific pathways, with all copies of a pathway on a particular subgenome becoming dominant alongside one another due to their specific affinities (Alger & Edgar, 2020; Wang et al., 2023). By extension of our discovery of gene level dominance, it may be possible to link pathways together and determine if this trend is observable. Chromosome degradation of the non-dominant subgenome is also a hallmark of dominance. When comparing subgenomes of *M. javanica* it is identified that chromosomes of subgenome B are shorter and more numerous than those of subgenomeA, and we suggest that subgenome B has undergone several chromosome fission events. This is suggestive that subgenome A is dominant. Due to the fragmented nature of the other assemblies phased and analysed in Chapter 3, we are unable to investigate if the same pattern is occurring in *M. arenaria*, *M. incognita*, and *M. luci*. Independent assembly of these species to a chromosome-scale would enable us to compare karyotypes and follow this route of investigation.

Knowing which of the subgenomes has lost a copy in the triploid species could also inform us of potential subgenome dominance. The end result of chromosome degradation,

fission, and subgenome dominance is the loss of entire copies and transformation to lower ploidy states, something which apparently has happened at least twice in the lineages analysed here. If it were shown that all triploid species have two copies of subgenome A, and one fissioned copy of subgenome B, it would further lend credence to subgenome A being dominant in most cases. Through the application of mapping approaches it may be possible to determine which subgenome of the triploid species - *M. luci, M. incognita,* and potentially *M. floridensis* - has lost a copy. Mapping of read libraries from these triploid species to a phased tetraploid assembly should quickly reveal a disparity of coverage on one of the subgenomes that correlates roughly to a 2:1 ratio, indicating that one of the subgenomes is present in only one copy. This would be especially interesting with *M. floridensis* which has been observed to be diploid in some lineages, suggesting a completion of the diploidisation process in some cases. Whether or not these diploid populations contained homoeologous copies (AB) or not (AA or BB) could inform us of the effects of subgenome dominance and diploidisation processes on the rest of the group. Why exactly *M. incognita* and *M. luci* became triploid, whilst *M. arenaria* and *M. javanica* remained tetraploid is unclear.

5.2.7 Methylation

Our analysis detected signals of CpG nucleotide methylation in the genome of *M. javanica* at levels above the error rate expected from the technologies used, in contrast to previous studies (Perfus-Barbeoch et al., 2014; Pratx et al., 2018). Novel allopolyploids undergo a period of genomic shock: an inhibition of TE control mechanisms leading to the mobilisation of TEs, facilitating large scale genomic alterations (McClintock, 1984; de Tomás & Vicient, 2023). This is thought to be managed through gradual TE methylation post-hybridisation (McClintock, 1984; de Tomás & Vicient, 2023). Alongside coping with genomic shock, methylation is thought to be one of the primary ways by which novel allopolyploids balance an increased gene dosage (Shan et al., 2024). RKN have been shown to be able to induce CpG methylation, albeit in the host (Hewezi, 2020; Bennett et al., 2022), as well as potentially contain genes orthologous to widely known methyltransferases (Perfus-Barbeoch et al., 2014). If RKN genomes do not exhibit methylation, the ability to cope with allopolyploidisation and resulting genomic shock would be either greatly reduced, or require a previously unidentified mechanism.

The methylation results presented here are interesting and would have benefitted from further investigation, but this was unfortunately beyond the scope of this project, however

we provide a foundation for investigation of methylation in MIG species, something that has previously not been thought to exist, or had only been suggested (Perfus-Barbeoch et al., 2014; Pratx et al., 2018). An overall comparison of methylation rates between subgenomes could perhaps show similar patterns seen under subgenome dominance, with one subgenome considerably more methylated (Bird et al., 2021; Yu et al., 2024). A finer scale comparison could be done at transcriptionally active sites, where it may be possible to see a pattern similar to that in Chapter 3.3.8 (Figure 3.11), where one gene is transcriptionally active and the other is silenced by methylation, in an alternating pattern along the genome.

Due to the close evolutionary history of VW4 and VW5 isolates we would expect a similar pattern to be observed in VW5. Oxford Nanopore data of VW5 has been generated which could provide answers to this question, however due to time and computational constraints, analysis of it was not possible within this thesis. Without assessing methylation in VW5 and comparing it to VW4 we cannot rule out that there is some sort of epigenetic difference between the lineages that is contributing to the differences we observe in phenotype, fitness and virulence.

5.2.8 VW5 gain-of-virulence

Using novel bioinformatic methods and workflows, we have identified a previously undiscovered deletion in a lineage of *M. javanica* (VW5) able to evade *Mi-1* induced host resistance. Furthermore, we identified several candidate genes for this gain of virulence within the deleted region, some of which are completely absent from VW5, and have predicted functions strongly indicative of a link to host-pathogen interactions. Allopolyploidy is thought to facilitate adaptation through the gain of novel functions and modifiable dosages through gene fractionation (Schiavinato et al., 2021). We show in Chapter 3 how gene fractionation appears to have occurred across the subgenomes of allotetraploid MIG species. In Chapter 4 we present what seems to be an adaptation through total gene loss, whereby a single functional copy is deleted. Loss of this effector reduces fitness on non-resistant hosts but allows infection of formerly resistant hosts containing the resistance gene *Mi-1* (Siddique, pers comms). This is an example of gene loss leading to adaptation (Sharma et al., 2018; Helsen et al., 2020). It could be that alleles of this gene have already deteriorated in the genome of *M. javanica* leaving only one functional copy, or that perhaps this gene was completely lost in one of the parent species and present in only one functional copy in the other parent, however coverage analysis of this region in VW4 shows three copies, with one being functional, suggesting that a region of one subgenome has been

completely lost sometime since hybridisation, possibly as a result of the copy containing the functional allele, on subgenome A, being dominant. Additional investigation of the differences between VW4 and VW5 and their impact on phenotype would benefit from analysis in the light of the haplotype-resolved *M. javanica* assembly (Dai et al., 2023), although care must be taken to account for possible differences between the isolate used for that assembly, and VW4. Assembling VW5 using this new assembly as a reference would enable VW5 to be haplotype-resolved, which in turn would increase the clarity of copy number and structural differences identified here, and make regions where genes are completely deleted in VW5 even more apparent.

PCR assays have been performed to detect the PLCP PB.2396 in *M. javanica*. These assays reliably detect the PLCP in VW4 and fail to detect it in VW5 (Siddique, pers comms). It follows that this method could be applied to other species of *Meloidogyne* where there are known to be avirulent and virulent isolates. The absence of PB.2396 in other virulent lineages could add weight to the evidence that the loss of this gene is the driver of *Mi-1* resistance evasion without requiring expensive and complex bioinformatic analysis. In addition, if it is discovered that the majority of *Mi-1* evasive virulent lineages are a result of deletion of this gene, a PCR assay to detect virulence more quickly than greenhouse experiments can be developed relatively easily.

The next stage of testing this gene for a link to gain-of-virulence would be to perform knockout experiments such as those performed by Gleason et al (2008). It may be that this gene is not the defining factor for virulence, and just that it resides in a pathway that when disrupted enables the RKN to evade *Mi-1,* so conclusively characterising this gene and determining its true function and position in any functional pathways is vital to a clear understanding of this gain-of-virulence mechanism.

5.3 Future directions and unanswered questions

Several interesting results described above were identified but unable to be investigated. Here we will address these findings, place them in the context of the current literature, and propose routes of further investigation.

5.3.1 Telomeres

Canonical telomeric repeats at chromosome termini were not found in our assembly, nor did we identify telomerase orthologs in our transcriptional sequencing or annotation (Chapter 2). Other studies have identified species specific composite repeats enriched at the contig termini of MIG species assemblies, determined to indeed be telomeric through fluorescence in-situ hybridisation (Mota et al., 2024). This means that *M. javanica* does not use usual methods for chromosome cap protection.

Several non-canonical methods of telomere composition and maintenance have been identified in other species, some exclusive of the use of telomerase. One example of this is the silkworm, for which telomeres contain many non-LTR retrotransposons and telomerase exhibits little to no enzymatic activity. Another more widely known example is that of dipterans, where telomerase is yet to be identified (Sasaki & Fujiwara, 2000; Mason et al., 2016). Dipteran telomeres consist of long blocks of complex repeat units, and are maintained and lengthened in some genera by homologous recombination (Cohn & Edström, 1992; Mason et al., 2011). It could be that telomeres in *Meloidogyne* consist of the composite motifs identified by Mota et al. (2024) and are maintained through a mechanism of homologous or homoeologous mitotic recombination similar to that in Diptera.

Further research could attempt to identify these repeats in our assembly of *M. javanica*, firstly in order to confirm that contigs in that assembly are chromosome-scale, and to also determine if there is interspecific variation of the repeat sequences used. It should be possible to search for telomerase or similar orthologs in previously published assemblies and read libraries of *Meloidogyne* species, as well as telomeric repeat sequences similar to those found by Mota et al. (2024), in order to determine how common this non-canonical method of telomeric preservation is across the group, or if it is restricted to the allopolyploid clade I species, or even just the MIG.

5.3.2 Parents of the MIG

The identification of potentially extant parent species to the allopolyploid RKN studied here is beyond the scope of this project, however we have provided new evidence that *M. luci* arose from a similar hybridisation event between the same parent species as the MIG. To gain further insight into the evolutionary history and potential origin of these subgenomes, future studies can utilise the advancements in genome assembly to accurately assemble polyploids, phase subgenomes, and generate subgenome scale phylogenies containing all

species in clade I. Based on phylogenies by Alvarez-Ortega et al. (2017), *M. luci* arises basal to the MIG, as well as several other species, including *M. phaseoli, M. arabicida,* and *M. konaensis.* This would mean that these species also have hybrid polyploid origin. They have chromosome numbers very similar to those of polyploid MIG species, and they are also mitotic parthenogens.

There are tests available that could detect similarities with other *Meloidogyne* species in order to detect signals of homology, and attempt to infer parentage. Sequence similarity is one option, comparing *Meloidogyne* assemblies to pan-subgenome assemblies of the allopolyploid species. Another option would be using the same ancestral k-mers detected in Chapter 2. New methods have been developed that can generate phylogenies from k-mer spectra (Tang et al., 2023; Van Etten et al., 2023). It may be possible to build a phylogeny from the ancestral repetitive k-mers of the *Meloidogyne* genus, including clade I subgenomes as independent taxa.

5.3.3 Homoeologous recombination

It may be possible to look for evidence of homeologous recombination in the annotations generated in Chapters 2 and 3, using a method similar to Szitenberg et al. (2017). CDS from either subgenome can be linked through pairwise alignment genome-wide, and measures of similarity calculated. A very high amount of sequence similarity between subgenomes of one species in a single orthogroup would indicate a site of homeologous conversion. Another method could be to use spectra of ancestral repetitive k-mers similarly to how assembly scaffolds were phased in Chapter 2. It is expected that these k-mers are exclusive to one or other subgenome, which means that areas where they are found on the opposing subgenome would suggest an area that had been converted.

5.3.4 Subgenome dominance

Investigations of selection and its effect on subgenome dominance are intensive and difficult (Qian et al., 2014; Pavlidis & Alachiotis, 2017; Alger & Edger, 2020; Conover & Wendel, 2022), and it is our opinion that although the analysis and results presented here addressing these phenomena are robust and indicative of a relative balance of dominance between subgenomes of most species, much more can be done to study this in detail. Several investigations of subgenome dominance, particularly in *Brassica* species, aim to

detect differences in expression of homoeologous loci, under the presumption that the dominant allele will exhibit higher expression rates than its homoeolog (Cheng et al., 2012; Bottani et al., 2018; Alger & Edger, 2020). The generation and analysis of quantifiable expression data of the species studied here could provide deeper insights into which subgenome is more dominant, as well as revealing differences in frequency of subfunctionalisation and shared dosage. It could be hypothesised that these will be more frequent in tetraploid species than triploids, driving retention of all original copies and impeding transformation to a lower ploidy state.

5.3.5 Investigating *Mi-1* resistance evasion mechanisms

Several species of *Meloidogyne* have been found to be able to evade host resistance bestowed by the *Mi-1* resistance gene (Ornat et al., 2001; Gleason, 2003; Tzortzakakis et al., 2005; Hajihassani et al., 2022; Ploeg et al., 2023). Given the results of Chapter 4, it can be hypothesised that deletion or reduced dosage of one or more transcriptionally active genes within the VW5 deletion is what enables phenotypic change and gain-of-virulence. This hypothesis is testable in several ways. First, the *copysnake* workflow described in Chapter 4 can be applied to sequencing libraries of other species with virulent lineages. Discrepancy between isolates similar those seen between VW4 and VW5 can reveal loci absent from the virulent isolate, suggesting deletion or disruption of a similar effector. Furthermore, depending on the release of phased chromosome-scale assemblies of MIG species such as those by Dai et al. (2023), a similar method of identifying deletions to those described in Chapter 4 can be applied. Genes within these potentially deleted regions can be compared across the group, with genes absent or disrupted in all virulent lineages gaining more weight as candidates for gain-of-virulence. In the absence of phased chromosome-scale assemblies or sequencing data of virulent lineages, PCR assays could be performed to detect if PB.2697.1 is present or absent. If absent, the implication that PB.2697.1 is involved in *Mi-1* evasion would increase. Similarly, knockout experiments of this gene in virulent lineages of other species could ascertain whether disruption of this PLCP also induces resistance evasion. If detected as present from PCR, it may mean that either absence of PB.2697.1 is not driving gain-of-virulence, or that PB.2697.1 is integral to a pathway that produces an effector that triggers *Mi-1* resistance. In this case, deletion of PB.2697.1 would disrupt the pathway and prevent production of the effector. This would generate a need to investigate this pathway, and determine if it is also disrupted in *Mi-1* evasive lineages.

5.4 Reproducibility

5.4.1 Bioinformatic resources

Throughout the performance of the research discussed here, many aspects of the bioinformatic methods were performed by reproducible workflows and scripts. These bioinformatic resources are also an important output of this work, and have been made publicly and freely available and accessible, either through Zenodo or Github. The largest of these resources is the *asmapp* workflow (Winter, 2022), used in Chapter 2 and Chapter 3 to appraise genome assemblies and generate descriptive statistics. This workflow is built in *snakemake*, employs the use of several bioinformatics packages alongside scripts written in *BASH*, *R*, and *Python*. *Asmapp* is currently available on Github for use. Another large workflow output from this project is *copysnake*. Built in *snakemake*, *copysnake* searches read libraries for reads containing a query sequence, then builds phylogenies of those detected sequences. It is our intention and hope that the public availability of these workflows will provide accessible and streamlined analysis tools to future researchers. The primary benefit of automated bioinformatic *snakemake* workflows, aside from the ability to version control and standardise analysis to achieve higher reproducibility, is their modularity. Being built in *snakemake*, the addition of further analysis and improved releases of packages contained within it, the workflow can adjust to computational progression in the field. One such example is in the use of k-mer spectra to assess completeness, a growing trend in genome assembly (Mapleson et al., 2017; Rhie et al., 2020). Addition of such packages is quick and straightforward, and their results can be incorporated into a final automated report very easily.

Alongside the larger workflows, almost all bioinformatic methods are maintained in *Jupyter* notebooks, made available on Zenodo for reference (10.5281/zenodo.13819447). This is not ideal, but several factors can limit the viability of an analysis being incorporated or performed in an automated workflow. Some analysis requires too much in-progress, researcher-led modification to be translated into an automated workflow. Similarly, the use cases of some such workflows would be very limited, making the time spent developing the workflow inefficient. *Jupyter* notebook based analyses can be as reproducible as those performed in automated workflows, but in absence of automation are susceptible to operator error.

Version control of all methods was managed by *conda* and all necessary files to rebuild and replicate *conda* environments are available alongside the appropriate resource. This

applies to all analyses described here, whether performed in automated workflows, notebooks, or standalone scripts. Every bioinformatic method described in this thesis is reproducible, with version-controlled methods and code publicly available.

5.5 Concluding remarks

In summary, this thesis has generated the most complete and contiguous assembly of the VW4 lineage of *Meloidogyne javanica* currently available, and annotated genes both structurally and functionally. We have determined that large amounts of synteny and collinearity remain both between homoeologous subgenomes, and between orthologous subgenomes across species. We have identified that *M. luci* subgenomes are orthologous with the subgenomes of the MIG species, indicating that all of these species share an origin and likely originated from the same hybridisation event between the same parental species. Extrapolating from this, we hypothesise that the vast majority of clade I *Meloidogyne* species - at least those more derived than *M. luci* - are descendants of hybridisation between the same parental species. We determine that subgenome dominance is acting with opposing pressures at a gene level, but that no single subgenome is overall dominant over the other. Alongside this we detect signals of methylation of the genome of *M. javanica*. Finally, we characterise and describe the *Cg-1* region using modern resources, discovering a large deletion in one copy of the virulent VW5 lineage of *M. javanica* responsible for the deletion or disruption of several transcriptionally active genes, some of which are heavily implicated in host-pathogen interactions and make very promising candidates for being responsible for gain-of-virulence adaptations in this species. Overall, we have contributed valid and consequential findings to this field of research, as well as generating several useful genomic resources, bioinformatic methods, workflows, and hypotheses for continued study and advancement toward our understanding of the genomics of this important and complex group of species.

6.0 References

Abad, P., Favery, B., Rosso, M.-N. & Castagnone-Sereno, P. (2003) Root-knot nematode parasitism and host response: molecular basis of a sophisticated interaction: Root-knot nematode parasitism and host response. *Molecular Plant Pathology*, 4(4), p. 217–224. Available online: https://doi.org/10.1046/j.1364-3703.2003.00170.x.

Abad, P., Gouzy, J., Aury, J.-M., Castagnone-Sereno, P., Danchin, E.G.J., Deleury, E., Perfus-Barbeoch, L., Anthouard, V., Artiguenave, F., Blok, V.C., Caillaud, M.-C., Coutinho, P.M., Dasilva, C., De Luca, F., Deau, F., Esquibet, M., Flutre, T., Goldstone, J.V., Hamamouch, N., Hewezi, T., Jaillon, O., Jubin, C., Leonetti, P., Magliano, M., Maier, T.R., Markov, G.V., McVeigh, P., Pesole, G., Poulain, J., Robinson-Rechavi, M., Sallet, E., Ségurens, B., Steinbach, D., Tytgat, T., Ugarte, E., van Ghelder, C., Veronico, P., Baum, T.J., Blaxter, M., Bleve-Zacheo, T., Davis, E.L., Ewbank, J.J., Favery, B., Grenier, E., Henrissat, B., Jones, J.T., Laudet, V., Maule, A.G., Quesneville, H., Rosso, M.-N., Schiex, T., Smant, G., Weissenbach, J. & Wincker, P. (2008) Genome sequence of the metazoan plant-parasitic nematode Meloidogyne incognita. *Nature Biotechnology*, 26(8), p. 909–915. Available online: https://doi.org/10.1038/nbt.1482.

Abolafia, J. & Shokoohi, E. (2017) Description of Stegelletina lingulata sp. n. (Nematoda, Rhabditida, Cephalobidae) from xeric environments in Iran. *Zootaxa*, 4358(3), p. 462–470. Available online: https://doi.org/10.11646/zootaxa.4358.3.4.

Abrantes, I., Almeida, M.T., Conceição, I.L., Esteves, I. & Maleita, C. (2023) Nematodes of potato and their management. In *Potato Production Worldwide*. Elsevier, p. 213–240. Available online: https://doi.org/10.1016/b978-0-12-822925-5.00024-4.

Acharya, K., Yan, G. & Plaisance, A. (2021) Effects of cover crops on population reduction of soybean cyst nematode (Heterodera glycines). *Plant Disease*, 105(4), p. 764–769. Available online: https://doi.org/10.1094/PDIS-08-20-1778-RE.

Adams, B.J., Dillman, A.R., Finlinson, C. & Others (2009) Molecular taxonomy and phylogeny. *Root-Knot Nematodes*, p. 119–138. Available online: https://books.google.co.uk/books?hl=en&lr=&id=UN3uHMr_UCoC&oi=fnd&pg=PA119&dq=adams+phylo+nematode&ots=_x_R4F1BfG&sig=jRv5bEaAS3YFdyy_FX4f4FdVNKo.

Ahmed, M., Roberts, N.G., Adediran, F., Smythe, A.B., Kocot, K.M. & Holovachov, O. (2022) Phylogenomic Analysis of the Phylum Nematoda: Conflicts and Congruences With Morphology, 18S rRNA, and Mitogenomes. *Frontiers in Ecology and Evolution*, 9. Available online: https://doi.org/10.3389/fevo.2021.769565.

Al Abadiyah Ralmi, N.H., Khandaker, M.M. & Mat, N. (2016) Occurrence and control of root knot nematode in crops: A review. *Australian Journal of Crop Science*, 10, p. 1649–1654. Available online: https://doi.org/10.21475/AJCS.2016.10.12.P7444.

Alger, E.I. & Edger, P.P. (2020) One subgenome to rule them all: underlying mechanisms of subgenome dominance. *Current Opinion in Plant Biology*, 54, p. 108–113. Available online: https://doi.org/10.1016/j.pbi.2020.03.004.

Ali, A., Thorgaard, G.H. & Salem, M. (2021) PacBio Iso-Seq Improves the Rainbow Trout Genome Annotation and Identifies Alternative Splicing Associated With Economically Important Phenotypes. *Frontiers in Genetics*, 12, p. 683408. Available online: https://doi.org/10.3389/fgene.2021.683408.

Allendorf, F.W., Hohenlohe, P.A. & Luikart, G. (2010) Genomics and the future of conservation genetics. *Nature Reviews. Genetics*, 11(10), p. 697–709. Available online: https://doi.org/10.1038/nrg2844.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), p. 403–410. Available online: https://doi.org/10.1016/S0022-2836(05)80360-2.

Álvarez-Ortega, S., Brito, J.A. & Subbotin, S.A. (2019) Multigene phylogeny of root-knot nematodes and molecular characterization of Meloidogyne nataliei Golden, Rose & Bird, 1981 (Nematoda: Tylenchida). *Scientific Reports*, 9(1), p. 11788. Available online: https://doi.org/10.1038/s41598-019-48195-0.

Anatskaya, O.V. & Vinogradov, A.E. (2022) Polyploidy as a fundamental phenomenon in evolution, development, adaptation and diseases. *International Journal of Molecular Sciences*, 23(7), p. 3542. Available online: https://doi.org/10.3390/ijms23073542.

Arita, M., Karsch-Mizrachi, I. & Cochrane, G. (2021) The international nucleotide sequence database collaboration. *Nucleic Acids Research*, 49(D1), p. D121–D124. Available online: https://doi.org/10.1093/nar/gkaa967.

Asamizu, E., Shirasawa, K., Hirakawa, H. & Iwahori, H. (2020) Root-knot nematode genetic diversity associated with host compatibility to sweetpotato cultivars. Molecular Plant Pathology, 21(8), p. 1088–1098. Available online: https://doi.org/10.1111/mpp.12961.

Balloux, F., Lehmann, L. & de Meeûs, T. (2003) The Population Genetics of Clonal and Partially Clonal Diploids. *Genetics*, 164(4), p. 1635–1644. Available online: https://doi.org/10.1093/genetics/164.4.1635.

Barbary, A., Djian-Caporalino, C., Palloix, A. & Castagnone-Sereno, P. (2015) Host genetic resistance to root-knot nematodes, *Meloidogyne* spp., in Solanaceae: from genes to the field: Host genetic resistance to root-knot nematodesin Solanaceae. *Pest Management Science*, 71(12), p. 1591–1598. Available online: https://doi.org/10.1002/ps.4091.

Barker, K.R., Carter, C.C. & Sasser, J.N. (1985) *An Advanced Treatise on Meloidogyne: Biology and control*. Department of Plant Pathology, North Carolina State University. Available online: https://play.google.com/store/books/details?id=ONo4AQAAIAAJ.

Barker, M.S., Arrigo, N., Baniaga, A.E., Li, Z. & Levin, D.A. (2015) On the relative abundance of autopolyploids and allopolyploids. *The New Phytologist* [Preprint]. Available online: https://doi.org/10.1111/nph.13698.

Bar-On, Y.M., Phillips, R. & Milo, R. (2018) The biomass distribution on Earth. *Proceedings of the National Academy of Sciences of the United States of America*, 115(25), p. 6506–6511. Available online: https://doi.org/10.1073/pnas.1711842115.

Barry, J.D. & McCulloch, R. (2001) Antigenic variation in trypanosomes: enhanced phenotypic variation in a eukaryotic parasite. *Advances in Parasitology*, 49, p. 1–70. Available online: https://doi.org/10.1016/s0065-308x(01)49037-3.

Baykal, P.I., Łabaj, P.P., Markowetz, F., Schriml, L.M., Stekhoven, D.J., Mangul, S. & Beerenwinkel, N. (2024) Genomic reproducibility in the bioinformatics era. *Genome Biology*, 25(1), p. 213. Available online: https://doi.org/10.1186/s13059-024-03343-2.

Bebber, D.P., Holmes, T. & Gurr, S.J. (2014) The global spread of crop pests and pathogens. *Global Ecology and Biogeography: A Journal of Macroecology*, 23(12), p. 1398–1407. Available online: https://doi.org/10.1111/geb.12214.

Bebber, D.P., Ramotowski, M.A.T. & Gurr, S.J. (2013) Crop pests and pathogens move polewards in a warming world. *Nature Climate Change*, 3(11), p. 985–988. Available online: https://doi.org/10.1038/nclimate1990.

Beiki, H., Liu, H., Huang, J., Manchanda, N., Nonneman, D., Smith, T.P.L., Reecy, J.M. & Tuggle, C.K. (2019) Improved annotation of the domestic pig genome through integration of Iso-Seq and RNA-seq data. *BMC Genomics*, 20(1), p. 344. Available online: https://doi.org/10.1186/s12864-019-5709-y.

Bellafiore, S., Shen, Z., Rosso, M.-N., Abad, P., Shih, P. & Briggs, S.P. (2008) Direct identification of the Meloidogyne incognita secretome reveals proteins with host cell reprogramming potential. *PLoS Pathogens*, 4(10), p. e1000192. Available online: https://doi.org/10.1371/journal.ppat.1000192.

Bell, G. (2021) *The masterpiece of nature: The evolution and genetics of sexuality*. London, England: Routledge (Routledge Revivals). Available online: https://books.google.co.uk/books?hl=en&lr=&id=ZcXADwAAQBAJ&oi=fnd&pg=PT17&dq=the+masterpiece+of+nature&ots=GzZRSUYLVy&sig=TV4_J3ItT2neD48Of1jWI1F5VbA.

Bennett, M., Hawk, T.E., Lopes-Caitar, V.S., Adams, N., Rice, J.H. & Hewezi, T. (2022) Establishment and maintenance of DNA methylation in nematode feeding sites. *Frontiers in Plant Science*, 13, p. 1111623. Available online: https://doi.org/10.3389/fpls.2022.1111623.

Berger, V.W. & Zhou, Y. (2014) Kolmogorov-Smirnov Test: Overview. *Wiley StatsRef: Statistics Reference Online*. Chichester, UK: John Wiley & Sons, Ltd. Available online: https://doi.org/10.1002/9781118445112.stat06558.

Bernard, G.C., Egnin, M. & Bonsi, C. (2017) The impact of plant-parasitic nematodes on agriculture and methods of control. *Nematology-Concepts, Diagnosis and Control*, 10, p. 121–151. Available online: https://books.google.com/books?hl=en&lr=&id=MQWQDwAAQBAJ&oi=fnd&pg=PA121&ots=bDd5LAqbhk&sig=4PxsNeS8U8QQ6PsoKYkfXy2msA4.

Bernard, G.C., Egnin, M. & Bonsi, C. (2017) The impact of plant-parasitic nematodes on agriculture and methods of control. *Nematology-Concepts, Diagnosis and Control*, 10. Available online: https://books.google.co.uk/books?hl=en&lr=&id=MQWQDwAAQBAJ&oi=fnd&pg=PA121&dq=agricultural+impact+root-knot+nematodes&ots=bBkbCxpflc&sig=1tg90hY-8Y2Z7V_6F8KGljCENoM.

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M. & Stadler, P.F. (2013) MITOS: improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*, 69(2), p. 313–319. Available online: https://doi.org/10.1016/j.ympev.2012.08.023.

Besemer, J. & Borodovsky, M. (2005) GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Research*, 33(Web Server issue), p. W451–4. Available online: https://doi.org/10.1093/nar/gki487.

Bethony, J., Brooker, S., Albonico, M., Geiger, S.M., Loukas, A., Diemert, D. & Hotez, P.J. (2006) Soil-transmitted helminth infections: ascariasis, trichuriasis, and hookworm. *The Lancet*, 367(9521), p. 1521–1532. Available online: https://doi.org/10.1016/S0140-6736(06)68653-4.

Betts, A., Rafaluk, C. & King, K.C. (2016) Host and Parasite Evolution in a Tangled Bank. *Trends in Parasitology*, 32(11), p. 863–873. Available online: https://doi.org/10.1016/j.pt.2016.08.003.

Bhutkar, A., Russo, S., Smith, T.F. & Gelbart, W.M. (2006) Techniques for multi-genome synteny analysis to overcome assembly limitations. *Genome Informatics. International Conference on Genome Informatics*, 17(2), p. 152–161. Available online: https://doi.org/10.11234/gi1990.17.2_152.

Biosciences, P. (no date) *Pbipa: Improved Phased Assembler*. Github. Available online: https://github.com/PacificBiosciences/pbipa [Accessed 19/12/2023].

Birchler, J.A. & Veitia, R.A. (2010) The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. *The New Phytologist*, 186(1), p. 54–62. Available online: https://doi.org/10.1111/j.1469-8137.2009.03087.x.

Birchler, J.A. & Yang, H. (2022) The multiple fates of gene duplications: Deletion, hypofunctionalization, subfunctionalization, neofunctionalization, dosage balance constraints, and neutral variation. *The Plant Cell*, 34(7), p. 2466–2474. Available online: https://doi.org/10.1093/plcell/koac076.

Birchler, J.A., Yao, H., Chudalayandi, S., Vaiman, D. & Veitia, R.A. (2010) Heterosis. *The Plant Cell*, 22(7), p. 2105–2112. Available online: https://doi.org/10.1105/tpc.110.076133.

Bird, D.M., Opperman, C.H. & Williamson, V.M. (2009) Plant Infection by Root-Knot Nematode. In Berg, R.H. & Taylor, C.G. (eds) *Cell Biology of Plant Nematode Parasitism*. Berlin, Heidelberg: Springer Berlin Heidelberg, p. 1–13. Available online: https://doi.org/10.1007/978-3-540-85215-5_1.

Bird, D.M., Williamson, V.M., Abad, P., McCarter, J., Danchin, E.G.J., Castagnone-Sereno, P. & Opperman, C.H. (2009) The genomes of root-knot nematodes. *Annual Review of Phytopathology*, 47, p. 333–351. Available online: https://doi.org/10.1146/annurev-phyto-080508-081839.

Bird, K.A.A.D. (2022) *Subgenome Dominance and Genome Evolution in Allopolyploids*. search.proquest.com. Available online: https://search.proquest.com/openview/b9e8bb462f437188d4f741ea1c731f75/1?pq-origsite=gscholar&cbl=18750&diss=y.

Bird, K.A., Niederhuth, C.E., Ou, S., Gehan, M., Pires, J.C., Xiong, Z., VanBuren, R. & Edger, P.P. (2021) Replaying the evolutionary tape to investigate subgenome dominance in allopolyploid Brassica napus. *The New Phytologist*, 230(1), p. 354–371. Available online: https://doi.org/10.1111/nph.17137.

Bird, K.A., VanBuren, R., Puzey, J.R. & Edger, P.P. (2018) The causes and consequences of subgenome dominance in hybrids and recent polyploids. *The New Phytologist*, 220(1), p. 87–93. Available online: https://doi.org/10.1111/nph.15256.

Blackmore, M.S. & Charnov, E.L. (1989) Adaptive Variation in Environmental Sex Determination in a Nematode. *The American Naturalist*, 134(5), p. 817–823. Available online: https://doi.org/10.1086/285013.

Blanc-Mathieu, R., Perfus-Barbeoch, L., Aury, J.-M., Da Rocha, M., Gouzy, J., Sallet, E., Martin-Jimenez, C., Bailly-Bechet, M., Castagnone-Sereno, P., Flot, J.-F., Kozlowski, D.K., Cazareth, J., Couloux, A., Da Silva, C., Guy, J., Kim-Jo, Y.-J., Rancurel, C., Schiex, T., Abad, P., Wincker, P. & Danchin, E.G.J. (2017) Hybridization and polyploidy enable genomic plasticity without sex in the most devastating plant-parasitic nematodes. *PLoS Genetics*, 13(6), p. e1006777. Available online: https://doi.org/10.1371/journal.pgen.1006777.

Blaxter, M. & Koutsovoulos, G. (2015) The evolution of parasitism in Nematoda. *Parasitology*, 142 Suppl 1, p. S26–39. Available online: https://doi.org/10.1017/S0031182014000791.

Blaxter, M., Leech, C. & Lunt, D.H. (2024) A catalogue of chromosome counts for Phylum Nematoda. *Wellcome Open Research*, 9, p. 55. Available online: https://doi.org/10.12688/wellcomeopenres.20550.1.

Blaxter, M., Mann, J., Chapman, T., Thomas, F., Whitton, C., Floyd, R. & Abebe, E. (2005) Defining operational taxonomic units using DNA barcode data. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 360(1462), p. 1935–1943. Available online: https://doi.org/10.1098/rstb.2005.1725.

Bock, D.G., Cai, Z., Elphinstone, C., González-Segovia, E., Hirabayashi, K., Huang, K., Keais, G.L., Kim, A., Owens, G.L. & Rieseberg, L.H. (2023) Genomics of plant speciation. *Plant Communications*, 4(5), p. 100599. Available online: https://doi.org/10.1016/j.xplc.2023.100599.

Bomblies, K., Jones, G., Franklin, C., Zickler, D. & Kleckner, N. (2016) The challenge of evolving stable polyploidy: could an increase in “crossover interference distance” play a central role? *Chromosoma*, 125(2), p. 287–300. Available online: https://doi.org/10.1007/s00412-015-0571-4.

Borgonie, G., García-Moyano, A., Litthauer, D., Bert, W., Bester, A., van Heerden, E., Möller, C., Erasmus, M. & Onstott, T.C. (2011) Nematoda from the terrestrial deep subsurface of South Africa. *Nature*, 474(7349), p. 79–82. Available online: https://doi.org/10.1038/nature09974.

Borowiec, M.L. (2016) AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ*, 4, p. e1660. Available online: https://doi.org/10.7717/peerj.1660.

Bottani, S., Zabet, N.R., Wendel, J.F. & Veitia, R.A. (2018) Gene expression dominance in allopolyploids: Hypotheses and models. *Trends in Plant Science*, 23(5), p. 393–402. Available online: https://doi.org/10.1016/j.tplants.2018.01.002.

Braga, P.H.P., Hébert, K., Hudgins, E.J., Scott, E.R., Edwards, B.P.M., Sánchez Reyes, L.L., Grainger, M.J., Foroughirad, V., Hillemann, F., Binley, A.D., Brookson, C.B., Gaynor, K.M., Shafiei Sabet, S., Güncan, A., Weierbach, H., Gomes, D.G.E. & Crystal-Ornelas, R. (2023) Not just for programmers: How GitHub can accelerate collaborative and reproducible research in ecology and evolution. *Methods in Ecology and Evolution*, 14(6), p. 1364–1380. Available online: https://doi.org/10.1111/2041-210X.14108.

Bretagnolle, F. & Thompson, J.D. (1995) Gametes with the somatic chromosome number: mechanisms of their formation and role in the evolution of autopolyploid plants. *The New Phytologist*, 129(1), p. 1–22. Available online: https://doi.org/10.1111/j.1469-8137.1995.tb03005.x.

Brito, I. (2021) Examining horizontal gene transfer in microbial communities. *Nature Reviews. Microbiology*, 19, p. 442–453. Available online: https://doi.org/10.1038/s41579-021-00534-7.

Bybee, S.M., Kalkman, V.J., Erickson, R.J., Frandsen, P.B., Breinholt, J.W., Suvorov, A., Dijkstra, K.-D.B., Cordero-Rivera, A., Skevington, J.H., Abbott, J.C., Sanchez Herrera, M., Lemmon, A.R., Moriarty Lemmon, E. & Ware, J.L. (2021) Phylogeny and classification of Odonata using targeted genomics. *Molecular Phylogenetics and Evolution*, 160, p. 107115. Available online: https://doi.org/10.1016/j.ympev.2021.107115.

Campbell, J.W., Fulcher, M.R., Grewell, B.J. & Young, S.L. (2023) Climate and pest interactions pose a cross-landscape management challenge to soil and water conservation. *Journal of Soil and Water Conservation*, 78(2), p. 39A–44A. Available online: https://doi.org/10.2489/jswc.2023.1025A.

Campbell, M.S., Holt, C., Moore, B. & Yandell, M. (2014) Genome Annotation and Curation Using MAKER and MAKER-P. *Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al.]*, 48, p. 4.11.1–39. Available online: https://doi.org/10.1002/0471250953.bi0411s48.

Cao, Y., Zhao, K., Xu, J., Wu, L., Hao, F., Sun, M., Dong, J., Chao, G., Zhang, H., Gong, X., Chen, Y., Chen, C., Qian, W., Pires, J.C., Edger, P.P. & Xiong, Z. (2023) Genome balance and dosage effect drive allopolyploid formation in *Brassica*. *Proceedings of the National Academy of Sciences of the United States of America*, 120(14), p. e2217672120. Available online: https://doi.org/10.1073/pnas.2217672120.

Capella-Gutiérrez, S., Silla-Martínez, J.M. & Gabaldón, T. (2009) TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* , 25(15), p. 1972–1973. Available online: https://doi.org/10.1093/bioinformatics/btp348.

Carlton, P.M., Davis, R.E. & Ahmed, S. (2022) Nematode chromosomes. *Genetics* [Preprint]. Available online: https://doi.org/10.1093/genetics/iyac014.

Carneiro, R.M.D., Correa, V.R., Almeida, M.R.A., Gomes, A.C.M., Deimi, A.M., Castagnone-Sereno, P. & Karssen, G. (2014) Meloidogyne luci n. sp. (Nematoda: Meloidogynidae), a root-knot nematode parasitising different crops in Brazil, Chile and Iran. *Nematology: International Journal of Fundamental and Applied Nematological Research*, 16(3), p. 289–301. Available online: https://doi.org/10.1163/15685411-00002765.

Castagnone-Sereno, P. (1994) Genetics of Meloidogyne Virulence Against Resistance Genes from Solanaceous Crops. In Lamberti, F., De Giorgi, C., & Bird, D.M. (eds) *Advances in Molecular Plant Nematology*. Boston, MA: Springer US, p. 261–276. Available online: https://doi.org/10.1007/978-1-4757-9080-1_22.

Castagnone-Sereno, P., Bongiovanni, M. & Dalmasso, A. (1993) Stable virulence against the tomato resistance Mi gene in the parthenogenetic root-knot nematode Meloidogyne incognita. *Phytopathology*, 83(8), p. 803–805. Available online: https://hal.inrae.fr/hal-02714638 [Accessed 27/07/2023].

Castagnone-Sereno, P. & Danchin, E.G.J. (2014) Parasitic success without sex – the nematode experience. *Journal of Evolutionary Biology*, 27(7), p. 1323–1333. Available online: https://doi.org/10.1111/jeb.12337.

Castagnone-Sereno, P., Danchin, E.G.J., Perfus-Barbeoch, L. & Abad, P. (2013) Diversity and evolution of root-knot nematodes, genus Meloidogyne: new insights from the genomic era. *Annual Review of Phytopathology*, 51, p. 203–220. Available online: https://doi.org/10.1146/annurev-phyto-082712-102300.

Castagnone-Sereno, P., Mulet, K., Danchin, E.G.J., Koutsovoulos, G.D., Karaulic, M., Da Rocha, M., Bailly-Bechet, M., Pratx, L., Perfus-Barbeoch, L. & Abad, P. (2019) Gene copy number variations as signatures of adaptive evolution in the parthenogenetic, plant-parasitic nematode Meloidogyne incognita. *Molecular Ecology*, 26, p. 906. Available online: https://doi.org/10.1111/mec.15095.

Cerca, J., Petersen, B., Lazaro-Guevara, J.M., Rivera-Colón, A., Birkeland, S., Vizueta, J., Li, S., Li, Q., Loureiro, J., Kosawang, C., Díaz, P.J., Rivas-Torres, G., Fernández-Mazuecos, M., Vargas, P., McCauley, R.A., Petersen, G., Santos-Bay, L., Wales, N., Catchen, J.M., Machado, D., Nowak, M.D., Suh, A., Sinha, N.R., Nielsen, L.R., Seberg, O., Gilbert, M.T.P., Leebens-Mack, J.H., Rieseberg, L.H. & Martin, M.D. (2022) The genomic basis of the plant island syndrome in Darwin's giant daisies. *Nature Communications*, 13(1), p. 3729. Available online: https://doi.org/10.1038/s41467-022-31280-w.

Challis, R.J., Richards, E., Rajan, J., Cochrane, G. & Blaxter, M. (2019) BlobToolKit – interactive quality assessment of genome assemblies. *G3* , 10, p. 1361–1374. Available online: https://doi.org/10.1534/g3.119.400908.

Chavhan, P.B., Khan, L.A., Raut, P.A., Maske, D.K., Rahman, S., Podchalwar, K.S. & Siddiqui, M.F.M. (2008) Prevalence of nematode parasites of ruminants at Nagpur. *Veterinary World*, 1, p. 140–140. Available online: https://www.veterinaryworld.org/2008/May/Prevalence%20of%20Nematode%20parasites%20of.pdf.

Cheng, F., Wu, J., Cai, X., Liang, J., Freeling, M. & Wang, X. (2018) Gene retention, fractionation and subgenome differences in polyploid plants. *Nature Plants*, 4(5), p. 258–268. Available online: https://doi.org/10.1038/s41477-018-0136-7.

Cheng, F., Wu, J., Fang, L., Sun, S., Liu, B., Lin, K., Bonnema, G. & Wang, X. (2012) Biased gene fractionation and dominant gene expression among the subgenomes of Brassica rapa. *PloS One*, 7(5), p. e36442. Available online: https://doi.org/10.1371/journal.pone.0036442.

Cheng, H., Concepcion, G.T., Feng, X., Zhang, H. & Li, H. (2021) Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods*, 18(2), p. 170–175. Available online: https://doi.org/10.1038/s41592-020-01056-5.

Chen, J. & Caswell-Chen, E.P. (2004) Facultative Vivipary is a Life-History Trait in Caenorhabditis elegans. *Journal of Nematology*, 36(2), p. 107–113. Available online: https://www.ncbi.nlm.nih.gov/pubmed/19262794.

Chen, L., Li, C., Li, B., Zhou, X., Bai, Y., Zou, X., Zhou, Z., He, Q., Chen, B., Wang, M., Xue, Y., Jiang, Z., Feng, J., Zhou, T., Liu, Z. & Xu, P. (2023) Evolutionary divergence of subgenomes in common carp provides insights into speciation and allopolyploid success. *Fundamental Research* [Preprint]. Available online: https://doi.org/10.1016/j.fmre.2023.06.011.

Chen, M.-Y., Teng, W.-K., Zhao, L., Hu, C.-X., Zhou, Y.-K., Han, B.-P., Song, L.-R. & Shu, W.-S. (2021) Comparative genomics reveals insights into cyanobacterial evolution and habitat adaptation. *The ISME Journal*, 15(1), p. 211–227. Available online: https://doi.org/10.1038/s41396-020-00775-z.

Chen, S., Zhou, Y., Chen, Y. & Gu, J. (2018) Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* , 34(17), p. i884–i890. Available online: https://doi.org/10.1093/bioinformatics/bty560.

Chen, Z.J. (2010) Molecular mechanisms of polyploidy and hybrid vigor. *Trends in Plant Science*, 15(2), p. 57–71. Available online: https://doi.org/10.1016/j.tplants.2009.12.003.

Chen, Z.J., Sreedasyam, A., Ando, A., Song, Q., De Santiago, L.M., Hulse-Kemp, A.M., Ding, M., Ye, W., Kirkbride, R.C., Jenkins, J., Plott, C., Lovell, J., Lin, Y.-M., Vaughn, R., Liu, B., Simpson, S., Scheffler, B.E., Wen, L., Saski, C.A., Grover, C.E., Hu, G., Conover, J.L., Carlson, J.W., Shu, S., Boston, L.B., Williams, M., Peterson, D.G., McGee, K., Jones, D.C., Wendel, J.F., Stelly, D.M., Grimwood, J. & Schmutz, J. (2020) Genomic diversifications of five Gossypium allopolyploid species and their impact on cotton improvement. *Nature Genetics*, 52(5), p. 525–533. Available online: https://doi.org/10.1038/s41588-020-0614-5.

Chin, C.-S., Peluso, P., Sedlazeck, F.J., Nattestad, M., Concepcion, G.T., Clum, A., Dunn, C., O'Malley, R., Figueroa-Balderas, R., Morales-Cruz, A., Cramer, G.R., Delledonne, M., Luo, C., Ecker, J.R., Cantu, D., Rank, D.R. & Schatz, M.C. (2016) Phased diploid genome assembly with single-molecule real-time sequencing. *Nature Methods*, 13(12), p. 1050–1054. Available online: https://doi.org/10.1038/nmeth.4035.

Chitwood, B.G., Specht, A.W. & Havis, L. (1952) Root-knot nematodes: Iii. Effects of Meloidogyne incognita and m. Javanica on some peach rootstocks. *Plant and Soil*, 4(1), p. 77–95. Available online: https://www.jstor.org/stable/42931598.

Choleva, L., Janko, K., De Gelas, K., Bohlen, J., Šlechtová, V., Rábová, M. & Ráb, P. (2012) Synthesis of clonality and polyploidy in vertebrate animals by hybridization between two sexual species. *Evolution; International Journal of Organic Evolution*, 66(7), p. 2191–2203. Available online: https://doi.org/10.1111/j.1558-5646.2012.01589.x.

Cifuentes, M., Grandont, L., Moore, G., Chèvre, A.M. & Jenczewski, E. (2010) Genetic regulation of meiosis in polyploid species: new insights into an old question. *The New Phytologist*, 186(1), p. 29–36. Available online: http://www.jstor.org/stable/27797519.

Clo, J., Padilla-García, N. & Kolář, F. (2022) Polyploidization as an opportunistic mutation: The role of unreduced gametes formation and genetic drift in polyploid establishment. *Journal of Evolutionary Biology*, 35(8), p. 1099–1109. Available online: https://doi.org/10.1111/jeb.14055.

Cock, P.J.A., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B. & de Hoon, M.J.L. (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* , 25(11), p. 1422–1423. Available online: https://doi.org/10.1093/bioinformatics/btp163.

Cohn, M. & Edström, J.E. (1992) Telomere-associated repeats in Chironomus form discrete subfamilies generated by gene conversion. *Journal of Molecular Evolution*, 35(2), p. 114–122. Available online: https://doi.org/10.1007/bf00183222.

Comai, L. (2005) The advantages and disadvantages of being polyploid. *Nature Reviews. Genetics*, 6(11), p. 836–846. Available online: https://doi.org/10.1038/nrg1711.

Conover, J.L. & Wendel, J.F. (2022) Deleterious mutations accumulate faster in allopolyploid than diploid cotton (Gossypium) and unequally between subgenomes. *Molecular Biology and Evolution*, 39(2). Available online: https://doi.org/10.1093/molbev/msac024.

Cornet, L. & Baurain, D. (2022) Contamination detection in genomic data: more is not enough. *Genome Biology*, 23(1), p. 60. Available online: https://doi.org/10.1186/s13059-022-02619-9.

Czech, B., Munafò, M., Ciabrelli, F., Eastwood, E.L., Fabry, M.H., Kneuss, E. & Hannon, G.J. (2018) PiRNA-guided genome defense: From biogenesis to silencing. *Annual Review of Genetics*, 52(1), p. 131–157. Available online: https://doi.org/10.1146/annurev-genet-120417-031441.

Dai, D., Xie, C., Zhou, Y., Bo, D., Zhang, S., Mao, S., Liao, Y., Cui, S., Zhu, Z., Wang, X., Li, F., Peng, D., Zheng, J. & Sun, M. (2023) Unzipped chromosome-level genomes reveal allopolyploid nematode origin pattern as unreduced gamete hybridization. *Nature Communications*, 14(1), p. 7156. Available online: https://doi.org/10.1038/s41467-023-42700-w.

Dainat, J. & Hereñú, D. (2020) *NBISweden/AGAT: AGAT-v0.4.0*. Available online: https://doi.org/10.5281/zenodo.3877441.

Dalio, R.J.D., Herlihy, J., Oliveira, T.S., McDowell, J.M. & Machado, M. (2018) Effector Biology in Focus: A Primer for Computational Prediction and Functional Characterization. *Molecular Plant-Microbe Interactions: MPMI*, 31(1), p. 22–33. Available online: https://doi.org/10.1094/MPMI-07-17-0174-FI.

Danchin, E., Koutsovoulos, G., Elashry, A., da Rocha, M., Poullet, M., Kozlowski, D., Sallet, E., Perfus-Barbeoch, L., Martin-Jimenez, C., Frey, J.E., Ahrens, C. & Kiewnick, S. (2020) The polyploid genome of the mitotic parthenogenetic root-knot nematode Meloidogyne enterolobii. figshare. Available online: https://doi.org/10.6084/M9.FIGSHARE.C.5007182.V1.

Danchin, E. & Rancurel, C. (2023) Meloidogyne luci gene predictions. Recherche Data Gouv. Available online: https://doi.org/10.57745/4VTGEC.

David, K.T. (2022) Global gradients in the distribution of animal polyploids. *Proceedings of the National Academy of Sciences of the United States of America*, 119(48), p. e2214070119. Available online: https://doi.org/10.1073/pnas.2214070119.

Davis, E.L., Hussey, R.S. & Baum, T.J. (2004) Getting to the roots of parasitism by nematodes. *Trends in Parasitology*, 20(3), p. 134–141. Available online: https://doi.org/10.1016/j.pt.2004.01.005.

Deakin, J.E., Potter, S., O'Neill, R., Ruiz-Herrera, A., Cioffi, M.B., Eldridge, M.D.B., Fukui, K., Marshall Graves, J.A., Griffin, D., Grutzner, F., Kratochvíl, L., Miura, I., Rovatsos, M., Srikulnath, K., Wapstra, E. & Ezaz, T. (2019) Chromosomics: Bridging the gap between genomes and chromosomes. *Genes*, 10(8), p. 627. Available online: https://doi.org/10.3390/genes10080627.

Despot-Slade, E., Širca, S., Mravinac, B., Castagnone-Sereno, P., Plohl, M. & Meštrović, N. (2022) Satellitome analyses in nematodes illuminate complex species history and show conserved features in satellite DNAs. *BMC Biology*, 20(1), p. 259. Available online: https://doi.org/10.1186/s12915-022-01460-7.

Dida, F. & Yi, G. (2021) Empirical evaluation of methods for de novo genome assembly. *PeerJ. Computer Science*, 7(e636), p. e636. Available online: https://doi.org/10.7717/peerj-cs.636.

Diemert, D.J., Bethony, J.M. & Hotez, P.J. (2008) Hookworm vaccines. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, 46(2), p. 282–288. Available online: https://doi.org/10.1086/524070.

van Dijk, E.L., Naquin, D., Gorrichon, K., Jaszczyszyn, Y., Ouazahrou, R., Thermes, C. & Hernandez, C. (2023) Genomics in the long-read sequencing era. *Trends in Genetics: TIG*, 39(9), p. 649–671. Available online: https://doi.org/10.1016/j.tig.2023.04.006.

Di Tommaso, P., Chatzou, M., Floden, E.W., Barja, P.P., Palumbo, E. & Notredame, C. (2017) Nextflow enables reproducible computational workflows. *Nature Biotechnology*, 35(4), p. 316–319. Available online: https://doi.org/10.1038/nbt.3820.

Dodds, P.N., Rafiqi, M., Gan, P.H.P., Hardham, A.R., Jones, D.A. & Ellis, J.G. (2009) Effectors of biotrophic fungi and oomycetes: pathogenicity factors and triggers of host resistance. *The New Phytologist*, 183(4), p. 993–1000. Available online: https://doi.org/10.1111/j.1469-8137.2009.02922.x.

Dodsworth, S., Chase, M.W. & Leitch, A.R. (2016) Is post-polyploidization diploidization the key to the evolutionary success of angiosperms?. *Botanical Journal of the Linnean Society. Linnean Society of London*, 180(1), p. 1–5. Available online: https://doi.org/10.1111/boj.12357.

Duan, H., Jones, A.W., Hewitt, T., Mackenzie, A., Hu, Y., Sharp, A., Lewis, D., Mago, R., Upadhyaya, N.M., Rathjen, J.P., Stone, E.A., Schwessinger, B., Figueroa, M., Dodds, P.N., Periyannan, S. & Sperschneider, J. (2021) Identification and correction of phase switches with Hi-C data in the Nanopore and HiFi chromosome-scale assemblies of the dikaryotic leaf rust fungus*Puccinia triticina*. *bioRxiv*. bioRxiv. Available online: https://doi.org/10.1101/2021.04.28.441890.

Duan, T., Sicard, A., Glémin, S. & Lascoux, M. (2024) Separating phases of allopolyploid evolution with resynthesized and natural Capsella bursa-pastoris. *eLife*, 12, p. RP88398. Available online: https://doi.org/10.7554/eLife.88398.

Du, K., Stöck, M., Kneitz, S., Klopp, C., Woltering, J.M., Adolfi, M.C., Feron, R., Prokopov, D., Makunin, A., Kichigin, I., Schmidt, C., Fischer, P., Kuhl, H., Wuertz, S., Gessner, J., Kloas, W., Cabau, C., Iampietro, C., Parrinello, H., Tomlinson, C., Journot, L., Braasch, I., Trifonov, V., Warren, W.C., Meyer, A., Guiguen, Y. & Schartl, M. (2020) The sterlet sturgeon genome sequence and the mechanisms of segmental rediploidization. *Nature Ecology & Evolution*, 4(6), p. 841–852. Available online: https://doi.org/10.1038/s41559-020-1166-x.

Duran, C., Edwards, D. & Batley, J. (2009) Genetic maps and the use of synteny. *Methods in Molecular Biology* , 513, p. 41–55. Available online: https://doi.org/10.1007/978-1-59745-427-8_3.

Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S. & Aiden, E.L. (2016) Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Systems*, 3(1), p. 99–101. Available online: https://doi.org/10.1016/j.cels.2015.07.012.

Dutta, T.K. & Phani, V. (2023) The pervasive impact of global climate change on plant-nematode interaction continuum. *Frontiers in Plant Science*, 14, p. 1143889. Available online: https://doi.org/10.3389/fpls.2023.1143889.

Eaton, D.A.R. (2020) Toytree: A minimalist tree visualization and manipulation library for Python. *Methods in Ecology and Evolution / British Ecological Society*. Edited by M. Matschiner, 11(1), p. 187–191. Available online: https://doi.org/10.1111/2041-210X.13313.

Eddaoudi, M., Ammati, M. & Rammah, A. (1997) Identification of the resistance breaking populations of Meloidogyne on tomatoes in Morocco and their effect on new sources of resistance. *Fundamental and Applied Agriculture* [Preprint].

Edger, P.P., McKain, M.R., Bird, K.A. & VanBuren, R. (2018) Subgenome assignment in allopolyploids: challenges and future directions. *Current Opinion in Plant Biology*, 42, p. 76–80. Available online: https://doi.org/10.1016/j.pbi.2018.03.006.

Edger, P.P., Poorten, T.J., VanBuren, R., Hardigan, M.A., Colle, M., McKain, M.R., Smith, R.D., Teresi, S.J., Nelson, A.D.L., Wai, C.M., Alger, E.I., Bird, K.A., Yocca, A.E., Pumplin, N., Ou, S., Ben-Zvi, G., Brodt, A., Baruch, K., Swale, T., Shiue, L., Acharya, C.B., Cole, G.S., Mower, J.P., Childs, K.L., Jiang, N., Lyons, E., Freeling, M., Puzey, J.R. & Knapp, S.J. (2019) Origin and evolution of the octoploid strawberry genome. *Nature Genetics*, 51(3), p. 541–547. Available online: https://doi.org/10.1038/s41588-019-0356-4.

Eisenback, J.D. & Triantaphyllou, H.H. (1991) Root-knot nematodes: Meloidogyne species and races. *Manual of Agricultural Nematology*, 1, p. 191–274. Available online: https://www.researchgate.net/profile/Jonathan_Eisenback/publication/283548298_Root-Knot_Nematodes_Meloidogyne_Species_and_Races/links/563e67e308ae34e98c4d93c1/Root-Knot-Nematodes-Meloidogyne-Species-and-Races.pdf.

Elad, Y. & Pertot, I. (2014) Climate change impacts on plant pathogens and plant diseases. *Journal of Crop Improvement*, 28(1), p. 99–139. Available online: https://doi.org/10.1080/15427528.2014.865412.

Elling, A.A. (2013) Major emerging problems with minor meloidogyne species. *Phytopathology*, 103(11), p. 1092–1102. Available online: https://doi.org/10.1094/PHYTO-01-13-0019-RVW.

Elshishka, M., Mladenov, A., Lazarova, S. & Peneva, V. (2023) Terrestrial nematodes from the Maritime Antarctic. *Biodiversity Data Journal*, 11, p. e102057. Available online: https://doi.org/10.3897/BDJ.11.e102057.

Emery, M., Willis, M.M.S., Hao, Y., Barry, K., Oakgrove, K., Peng, Y., Schmutz, J., Lyons, E., Pires, J.C., Edger, P.P. & Conant, G.C. (2018) Preferential retention of genes from one parental genome after polyploidy illustrates the nature and scope of the genomic conflicts induced by hybridization. *PLoS Genetics*, 14(3), p. e1007267. Available online: https://doi.org/10.1371/journal.pgen.1007267.

Emms, D.M. & Kelly, S. (2018) OrthoFinder2: fast and accurate phylogenomic orthology analysis from gene sequences. Available online: https://doi.org/10.1101/466201.

Engelhardt, J., Scheer, O., Stadler, P.F. & Prohaska, S.J. (2022) Evolution of DNA methylation across Ecdysozoa. *Journal of Molecular Evolution*, 90(1), p. 56–72. Available online: https://doi.org/10.1007/s00239-021-10042-0.

Etienne-Manneville, S. & Hall, A. (2002) Rho GTPases in cell biology. *Nature*, 420(6916), p. 629–635. Available online: https://doi.org/10.1038/nature01148.

Favery, B., Dubreuil, G., Chen, M.-S., Giron, D. & Abad, P. (2020) Gall-Inducing Parasites: Convergent and Conserved Strategies of Plant Manipulation by Insects and Nematodes. *Annual Review of Phytopathology*, 58, p. 1–22. Available online: https://doi.org/10.1146/annurev-phyto-010820-012722.

Favery, B., Quentin, M., Jaubert-Possamai, S. & Abad, P. (2016) Gall-forming root-knot nematodes hijack key plant cellular functions to induce multinucleate and hypertrophied feeding cells. *Journal of Insect Physiology* [Preprint]. Available online: https://doi.org/10.1016/J.JINSPHYS.2015.07.013.

Feldman, M. & Levy, A.A. (2012) Genome evolution due to allopolyploidization in wheat. *Genetics*, 192(3), p. 763–774. Available online: https://doi.org/10.1534/genetics.112.146316.

Feyisa, B. (2021) Review on Root Knot Nematodes (Rkns): Impact and Methods For Control. *Journal of Plant Pathology & Microbiology* [Preprint]. Available online: https://doi.org/10.35248/2157-7471.21.12.547.

Forghani, F. & Hajihassani, A. (2020) Recent advances in the development of environmentally benign treatments to control root-knot nematodes. *Frontiers in Plant Science*, 11, p. 1125. Available online: https://doi.org/10.3389/fpls.2020.01125.

Formenti, G., Theissinger, K., Fernandes, C., Bista, I., Bombarely, A., Bleidorn, C., Ciofi, C., Crottini, A., Godoy, J.A., Höglund, J., Malukiewicz, J., Mouton, A., Oomen, R.A., Paez, S., Palsbøll, P.J., Pampoulie, C., Ruiz-López, M.J., Svardal, H., Theofanopoulou, C., de Vries, J., Waldvogel, A.-M., Zhang, G., Mazzoni, C.J., Jarvis, E.D., Bálint, M. & European Reference Genome Atlas (ERGA) Consortium (2022) The era of reference genomes in conservation genomics. *Trends in Ecology & Evolution*, 37(3), p. 197–202. Available online: https://doi.org/10.1016/j.tree.2021.11.008.

Fouché, S., Plissonneau, C. & Croll, D. (2018) The birth and death of effectors in rapidly evolving filamentous pathogen genomes. *Current Opinion in Microbiology*, 46, p. 34–42. Available online: https://doi.org/10.1016/j.mib.2018.01.020.

Fox, D.T., Soltis, D.E., Soltis, P.S., Ashman, T.-L. & Van de Peer, Y. (2020) Polyploidy: A biological force from cells to ecosystems. *Trends in Cell Biology*, 30(9), p. 688–694. Available online: https://doi.org/10.1016/j.tcb.2020.06.006.

Frawley, L.E. & Orr-Weaver, T.L. (2015) Polyploidy. *Current Biology: CB*, 25(9), p. R353–8. Available online: https://doi.org/10.1016/j.cub.2015.03.037.

Freckman, D.W., Mankau, R. & Ferris, H. (1975) Nematode community structure in desert soils: nematode recovery. *Journal of Nematology*, 7(4), p. 343–346. Available online: https://www.ncbi.nlm.nih.gov/pubmed/19308179.

Freckman, D.W., Mankau, R. & Sher, S.A. (1974) Biology of nematodes in desert ecosystems. Available online: https://digitalcommons.usu.edu/cgi/viewcontent.cgi?article=1074&context=dbiome_progress.

Fu, Y., Mahmoud, M., Muraliraman, V.V., Sedlazeck, F.J. & Treangen, T.J. (2021) Vulcan: Improved long-read mapping and structural variant calling via dual-mode alignment. *GigaScience*, 10(9). Available online: https://doi.org/10.1093/gigascience/giab063.

Gabriel, L., Hoff, K.J., Bruna, T. & Lomsadze, A. (2023) The BRAKER3 genome annotation pipeline. *Plant and Animal* [Preprint]. Available online: https://www.researchgate.net/profile/Lars-Gabriel-3/publication/367409637_The_BRAKER3_Genome_Annotation_Pipeline/links/63d14dbbd9fb5967c204c184/The-BRAKER3-Genome-Annotation-Pipeline.pdf.

Gan, X., Li, S., Zong, Y., Cao, D., Li, Y., Liu, R., Cheng, S., Liu, B. & Zhang, H. (2021) Chromosome-Level Genome Assembly Provides New Insights into Genome Evolution and Tuberous Root Formation of Potentilla anserina. *Genes*, 12(12). Available online: https://doi.org/10.3390/genes12121993.

Gao, F., Liu, X., Wu, X.-P., Wang, X.-L., Gong, D., Lu, H., Xia, Y., Song, Y., Wang, J., Du, J., Liu, S., Han, X., Tang, Y., Yang, H., Jin, Q., Zhang, X. & Liu, M. (2012) Differential DNA methylation in discrete developmental stages of the parasitic nematode Trichinella spiralis. *Genome Biology*, 13(10), p. R100. Available online: https://doi.org/10.1186/gb-2012-13-10-r100.

García, L.E. & Sánchez-Puerta, M.V. (2015) Comparative and evolutionary analyses of Meloidogyne spp. Based on mitochondrial genome sequences. *PloS One*, 10(3), p. e0121142. Available online: https://doi.org/10.1371/journal.pone.0121142.

Gauthier, T.D. (2001) Detecting Trends Using Spearman's Rank Correlation Coefficient. *Environmental Forensics*, 2(4), p. 359–362. Available online: https://doi.org/10.1006/enfo.2001.0061.

Gerič Stare, B., Strajnar, P., Susič, N., Urek, G. & Širca, S. (2017) Reported populations of Meloidogyne ethiopica in Europe identified as Meloidogyne luci. *Plant Disease*, 101(9), p. 1627–1632. Available online: https://doi.org/10.1094/PDIS-02-17-0220-RE.

Gerlach, S.A. & Riemann, F. (1973) The Bremerhaven Checklist of Aquatic Nematodes: A catalog of Nematoda Adenophorea excluding the Dorylaimida. Veroffentlichungen des Instituts fur Meeresforschung in Bremerhaven. Supplement.

Gibson, G. & Wagner, G. (2000) Canalization in evolutionary genetics: a stabilizing theory?. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 22(4), p. 372–380. Available online: https://doi.org/10.1002/(SICI)1521-1878(200004)22:4<372::AID-BIES7>3.0.CO;2-J.

Gilbert, S.F. (2000) Early Development of the Nematode Caenorhabditis elegans. In *Developmental Biology. 6th Edition*. Sinauer Associates. Available online: https://www.ncbi.nlm.nih.gov/books/NBK10011/ [Accessed 05/08/2024].

Gleason, C.A. (2003) *Comparison of two strains of Meloidogyne javanica differing in virulence on tomato with the resistance gene Mi, and identification of a polymorphism that correlates with avirulence*. search.proquest.com. Available online: https://search.proquest.com/openview/94c5810600be73264e1e424f3f7d8e07/1?pq-origsite=gscholar&cbl=18750&diss=y.

Gleason, C.A., Liu, Q.L. & Williamson, V.M. (2008) Silencing a candidate nematode effector gene corresponding to the tomato resistance gene Mi-1 leads to acquisition of virulence. *Molecular Plant-Microbe Interactions: MPMI*, 21(5), p. 576–585. Available online: https://doi.org/10.1094/MPMI-21-5-0576.

Glover, N.M., Redestig, H. & Dessimoz, C. (2016) Homoeologs: What Are They and How Do We Infer Them?. *Trends in Plant Science*, 21(7), p. 609–621. Available online: https://doi.org/10.1016/j.tplants.2016.02.005.

Goldman, N. & Yang, Z. (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular Biology and Evolution*, 11(5), p. 725–736. Available online: https://academic.oup.com/mbe/article-abstract/11/5/725/1008711.

Goldstein, P. & Triantaphyllou, A.C. (1982) The synaptonemal complexes of Meloidogyne: relationship of structure and evolution of parthenogenesis. *Chromosoma*, 87(1), p. 117–124. Available online: https://doi.org/10.1007/bf00333513.

Gross, S.M. & Williamson, V.M. (2011) Tm1: a mutator/foldback transposable element family in root-knot nematodes. *PloS One*, 6(9), p. e24534. Available online: https://doi.org/10.1371/journal.pone.0024534.

Guan, D., McCarthy, S.A., Wood, J., Howe, K., Wang, Y. & Durbin, R. (2020) Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* , 36(9), p. 2896–2898. Available online: https://doi.org/10.1093/bioinformatics/btaa025.

Gubanov, N.M. & Others (1951) A giant nematode from the placenta of cetaceans Placenttnema gigantissima ngn sp. *Doklady Akademii Nauk SSSR*, 77(6), p. 1123–1125. Available online: https://www.cabdirect.org/cabdirect/abstract/19510801338.

Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics* , 29(8), p. 1072–1075. Available online: https://doi.org/10.1093/bioinformatics/btt086.

Hajihassani, A., Marquez, J., Woldemeskel, M. & Hamidi, N. (2022) Identification of Four Populations of Meloidogyne incognita in Georgia, United States, Capable of Parasitizing Tomato-Bearing Mi-1.2 Gene. *Plant Disease*, 106(1), p. 137–143. Available online: https://doi.org/10.1094/PDIS-05-21-0902-RE.

Handoo, Z.A., Nyczepir, A.P., Esmenjaud, D., van der Beek, J.G., Castagnone-Sereno, P., Carta, L.K., Skantar, A.M. & Higgins, J.A. (2004) Morphological, Molecular, and Differential-Host Characterization of Meloidogyne floridensis n. sp. (Nematoda: Meloidogynidae), a Root-Knot Nematode Parasitizing Peach in Florida. *Journal of Nematology*, 36(1), p. 20–35. Available online: https://www.ncbi.nlm.nih.gov/pubmed/19262784.

Harris, S., Rudnicki, K.S. & Haber, J.E. (1993) Gene conversions and crossing over during homologous and homeologous ectopic recombination in Saccharomyces cerevisiae. *Genetics*, 135(1), p. 5–16. Available online: https://doi.org/10.1093/genetics/135.1.5.

Hartman, K.M. & Sasser, J.N. (1985) Identification of Meloidogyne species on the basis of differential host test and perineal pattern morphology. *An Advanced Treatise on Meloidogyne*, 2, p. 69–77. Available online: https://books.google.com/books?hl=en&lr=&id=gNo4AQAAIAAJ&oi=fnd&pg=PA69&dq=hartman+and+sasser&ots=r0T3kOF2TR&sig=qp5-qepxxIy-MaZ3Yxq8q3OTlBg.

Hemara, L.M., Jayaraman, J., Sutherland, P.W., Montefiori, M., Arshed, S., Chatterjee, A., Chen, R., Andersen, M.T., Mesarich, C.H., van der Linden, O., Yoon, M., Schipper, M.M., Vanneste, J.L., Brendolise, C. & Templeton, M.D. (2022) Effector loss drives adaptation of Pseudomonas syringae pv. actinidiae biovar 3 to Actinidia arguta. *PLoS Pathogens*, 18(5), p. e1010542. Available online: https://doi.org/10.1371/journal.ppat.1010542.

Heslop-Harrison, J., Schwarzacher, T. & Liu, Q. (2022) Polyploidy: its consequences and enabling role in plant diversification and evolution. *Annals of Botany*, 131, p. 1–10. Available online: https://doi.org/10.1093/aob/mcac132.

Hettmansperger, T.P. & McKean, J.W. (2010) *Robust Nonparametric Statistical Methods*. 2nd edition. Boca Raton, FL: CRC Press (Chapman & Hall/CRC Monographs on Statistics and Applied Probability). Available online: https://books.google.com/books?hl=en&lr=&id=6w3LBQAAQBAJ&oi=fnd&pg=PP1&dq=robust+nonparametric+statistical+methods&ots=r8X5m26QFH&sig=rT0s_KIEw2jD-QpeLlLu9KxQ9RI.

Hewezi, T. (2020) Epigenetic mechanisms in nematode-plant interactions. *Annual Review of Phytopathology*, 58(1), p. 119–138. Available online: https://doi.org/10.1146/annurev-phyto-010820-012805.

Hodda, M. (2022) Phylum Nematoda: a classification, catalogue and index of valid genera, with a census of valid species. *Zootaxa*, 5114(1), p. 1–289. Available online: https://doi.org/10.11646/zootaxa.5114.1.1.

Hodgkin, J. (1983) Two types of sex determination in a nematode. *Nature*, 304(5923), p. 267–268. Available online: https://doi.org/10.1038/304267a0.

Hoff, K.J. & Stanke, M. (2019) Predicting Genes in Single Genomes with AUGUSTUS. *Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al.]*, 65(1), p. e57. Available online: https://doi.org/10.1002/cpbi.57.

Hogger, C.H. & Bird, G.W. (1974) Weeds and covercrops as overwintering hosts of plant parasitic nematodes of soybean and cotton in Georgia. *Journal of Nematology*, 6(5), p. 142–143. Available online: https://www.cabidigitallibrary.org/doi/full/10.5555/19740816975.

Hollister, J.D. (2015) Polyploidy: adaptation to the genomic environment. *The New Phytologist*, 205(3), p. 1034–1039. Available online: https://doi.org/10.1111/nph.12939.

Holovachov, O. (2014) Nematodes from terrestrial and freshwater habitats in the Arctic. *Biodiversity Data Journal*, (2), p. e1165. Available online: https://doi.org/10.3897/BDJ.2.e1165.

Holt, C. & Yandell, M. (2011) MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*, 12, p. 491. Available online: https://doi.org/10.1186/1471-2105-12-491.

Hölzer, M. & Marz, M. (2019) De novo transcriptome assembly: A comprehensive cross-species comparison of short-read RNA-Seq assemblers. *GigaScience*, 8(5). Available online: https://doi.org/10.1093/gigascience/giz039.

van den Hoogen, J., Geisen, S., Routh, D., Ferris, H., Traunspurger, W., Wardle, D.A., de Goede, R.G.M., Adams, B.J., Ahmad, W., Andriuzzi, W.S., Bardgett, R.D., Bonkowski, M., Campos-Herrera, R., Cares, J.E., Caruso, T., de Brito Caixeta, L., Chen, X., Costa, S.R., Creamer, R., Mauro da Cunha Castro, J., Dam, M., Djigal, D., Escuer, M., Griffiths, B.S., Gutiérrez, C., Hohberg, K., Kalinkina, D., Kardol, P., Kergunteuil, A., Korthals, G., Krashevska, V., Kudrin, A.A., Li, Q., Liang, W., Magilton, M., Marais, M., Martín, J.A.R., Matveeva, E., Mayad, E.H., Mulder, C., Mullin, P., Neilson, R., Nguyen, T.A.D., Nielsen, U.N., Okada, H., Rius, J.E.P., Pan, K., Peneva, V., Pellissier, L., Carlos Pereira da Silva, J., Pitteloud, C., Powers, T.O., Powers, K., Quist, C.W., Rasmann, S., Moreno, S.S., Scheu, S., Setälä, H., Sushchuk, A., Tiunov, A.V., Trap, J., van der Putten, W., Vestergård, M., Villenave, C., Waeyenberge, L., Wall, D.H., Wilschut, R., Wright, D.G., Yang, J.-I. & Crowther, T.W. (2019) Soil nematode abundance and functional group composition at a global scale. *Nature*, 572(7768), p. 194–198. Available online: https://doi.org/10.1038/s41586-019-1418-6.

Howe, K., Chow, W., Collins, J., Pelan, S., Pointon, D.-L., Sims, Y., Torrance, J., Tracey, A. & Wood, J. (2021) Significantly improving the quality of genome assemblies through curation. *GigaScience*, 10(1), p. giaa153. Available online: https://doi.org/10.1093/gigascience/giaa153.

Howe, K.L., Bolt, B.J., Shafie, M., Kersey, P. & Berriman, M. (2017) WormBase ParaSite - a comprehensive resource for helminth genomics. *Molecular and Biochemical Parasitology*, 215, p. 2–10. Available online: https://doi.org/10.1016/j.molbiopara.2016.11.005.

Hudson, R. (2021) Should We Strive to Make Science Bias-Free? A Philosophical Assessment of the Reproducibility Crisis. *Journal for General Philosophy of Science. Zeitschrift Fur Allgemeine Wissenschaftstheorie*, 52(3), p. 389–405. Available online: https://doi.org/10.1007/s10838-020-09548-w.

Huerta-Cepas, J., Serra, F. & Bork, P. (2016) ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Molecular Biology and Evolution*, 33(6), p. 1635–1638. Available online: https://doi.org/10.1093/molbev/msw046.

Hugot, J.-P., Baujard, P. & Morand, S. (2001) Biodiversity in helminths and nematodes as a field of study: an overview. *Nematology: International Journal of Fundamental and Applied Nematological Research*, 3(3), p. 199–208. Available online: https://doi.org/10.1163/156854101750413270.

Hussain, M., Zouhar, M. & Ryšánek, P. (2016) Comparison between biological and chemical management of root knot nematode, Meloidogyne hapla. *Pakistan Journal of Zoology*, 49(1), p. 205–210. Available online: https://doi.org/10.17582/journal.pjz/2017.49.1.205.210.

Hussey, R.S. (1989) Disease-Inducing Secretions of Plant-Parasitic Nematodes. *Annual Review of Phytopathology*, 27(Volume 27, 1989), p. 123–141. Available online: https://doi.org/10.1146/annurev.py.27.090189.001011.

Ijaz, A.Z., Ali, R.H., Sarwar, A., Ali Khan, T. & Baig, M.M. (2022) Importance of Synteny in Homology Inference. In *2022 17th International Conference on Emerging Technologies (ICET)*. *2022 17th International Conference on Emerging Technologies (ICET)*, IEEE, p. 234–239. Available online: https://doi.org/10.1109/icet56601.2022.10004649.

Iriarte, A., Lamolle, G. & Musto, H. (2021) Codon Usage Bias: An Endless Tale. *Journal of Molecular Evolution*, 89(9-10), p. 589–593. Available online: https://doi.org/10.1007/s00239-021-10027-z.

Jagdale, S., Rao, U. & Giri, A.P. (2021) Effectors of Root-Knot Nematodes: An Arsenal for Successful Parasitism. *Frontiers in Plant Science*, 12, p. 800030. Available online: https://doi.org/10.3389/fpls.2021.800030.

Jain, M., Koren, S., Miga, K.H., Quick, J., Rand, A.C., Sasani, T.A., Tyson, J.R., Beggs, A.D., Dilthey, A.T., Fiddes, I.T., Malla, S., Marriott, H., Nieto, T., O'Grady, J., Olsen, H.E., Pedersen, B.S., Rhie, A., Richardson, H., Quinlan, A.R., Snutch, T.P., Tee, L., Paten, B., Phillippy, A.M., Simpson, J.T., Loman, N.J. & Loose, M. (2018) Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnology*, 36(4), p. 338–345. Available online: https://doi.org/10.1038/nbt.4060.

Janko, K., Pačes, J., Wilkinson-Herbots, H., Costa, R.J., Roslein, J., Drozd, P., Iakovenko, N., Rídl, J., Hroudová, M., Kočí, J., Reifová, R., Šlechtová, V. & Choleva, L. (2018) Hybrid asexuality as a primary postzygotic barrier between nascent species: On the interconnection between asexuality, hybridization and speciation. *Molecular Ecology*, 27(1), p. 248–263. Available online: https://doi.org/10.1111/mec.14377.

Janssen, T., Karssen, G., Topalović, O., Coyne, D. & Bert, W. (2017) Integrative taxonomy of root-knot nematodes reveals multiple independent origins of mitotic parthenogenesis. *PloS One*, 12(3), p. e0172190. Available online: https://doi.org/10.1371/journal.pone.0172190.

Janssen, T., Karssen, G., Verhaeven, M., Coyne, D. & Bert, W. (2016) Mitochondrial coding genome analysis of tropical root-knot nematodes (Meloidogyne) supports haplotype based diagnostics and reveals evidence of recent reticulate evolution. *Scientific Reports*, 6, p. 22591. Available online: https://doi.org/10.1038/srep22591.

Jaron, K., Jaron, K., Bast, J., Nowell, R., Nowell, R., Ranallo-Benavidez, T.R., Robinson-Rechavi, M., Robinson-Rechavi, M. & Schwander, T. (2018) Genomic Features of Parthenogenetic Animals. *The Journal of Heredity*, 112, p. 19–33. Available online: https://doi.org/10.1093/jhered/esaa031.

Jaron, K.S., Bast, J., Nowell, R.W., Ranallo-Benavidez, T.R., Robinson-Rechavi, M. & Schwander, T. (2021) Genomic Features of Parthenogenetic Animals. *The Journal of Heredity*, 112(1), p. 19–33. Available online: https://doi.org/10.1093/jhered/esaa031.

Jeffares, D.C., Tomiczek, B., Sojo, V. & dos Reis, M. (2015) A beginners guide to estimating the non-synonymous to synonymous rate ratio of all protein-coding genes in a genome. *Methods in Molecular Biology* , 1201, p. 65–90. Available online: https://doi.org/10.1007/978-1-4939-1438-8_4.

Jiang, N. (2013) Overview of repeat annotation and de novo repeat identification. *Methods in Molecular Biology (Clifton, N.J.)*, 1057, p. 275–287. Available online: https://doi.org/10.1007/978-1-62703-568-2_20.

Jones, J.D.G. & Dangl, J.L. (2006) The plant immune system. *Nature*, 444(7117), p. 323–329. Available online: https://doi.org/10.1038/nature05286.

Jones, J.T., Haegeman, A., Danchin, E., Gaur, H.S., Helder, J., Jones, M.G.K., Kikuchi, T., Manzanilla-López, R., Palomares-Rius, J.E., Wesemael, W. & Perry, R. (2013) Top 10 plant-parasitic nematodes in molecular plant pathology. *Molecular Plant Pathology*. Available online: https://doi.org/10.1111/MPP.12057.

Jones, J.T., Haegeman, A., Danchin, E.G.J., Gaur, H.S., Helder, J., Jones, M.G.K., Kikuchi, T., Manzanilla-López, R., Palomares-Rius, J.E., Wesemael, W.M.L. & Others (2013) Top 10 plant-parasitic nematodes in molecular plant pathology. *Molecular Plant Pathology*, 14(9), p. 946–961. Available online: https://bsppjournals.onlinelibrary.wiley.com/doi/abs/10.1111/mpp.12057.

Jones, J.T., Haegeman, A., Danchin, E.G.J., Gaur, H.S., Helder, J., Jones, M.G.K., Kikuchi, T., Manzanilla-López, R., Palomares-Rius, J.E., Wesemael, W.M.L. & Perry, R.N. (2013) Top 10 plant-parasitic nematodes in molecular plant pathology. *Molecular Plant Pathology*, 14(9), p. 946–961. Available online: https://doi.org/10.1111/mpp.12057.

Joseph, S., Mekete, T., Danquah, W.B. & Noling, J. (2016) First Report of Meloidogyne haplanaria Infecting Mi-Resistant Tomato Plants in Florida and Its Molecular Diagnosis Based on Mitochondrial Haplotype. *Plant Disease*, 100(7), p. 1438–1445. Available online: https://doi.org/10.1094/PDIS-09-15-1113-RE.

Jung, H., Ventura, T., Chung, J.S., Kim, W.-J., Nam, B.-H., Kong, H.J., Kim, Y.-O., Jeon, M.-S. & Eyun, S.-I. (2020) Twelve quick steps for genome assembly and annotation in the classroom. *PLOS Computational Biology*. Edited by F. Ouellette, 16(11), p. e1008325. Available online: https://doi.org/10.1371/journal.pcbi.100832510.1371/journal.pcbi.1008325.g00110.1371/journal.pcbi.1008325.t00110.1371/journal.pcbi.1008325.t00210.1371/journal.pcbi.1008325.t003.

Kao, R.H. (2007) Asexuality and the coexistence of cytotypes. *The New Phytologist*, 175(4), p. 764–772. Available online: https://doi.org/10.1111/j.1469-8137.2007.02145.x.

Katoh, K., Misawa, K., Kuma, K.-I. & Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30(14), p. 3059–3066. Available online: https://doi.org/10.1093/nar/gkf436.

Katz, K., Shutov, O., Lapoint, R., Kimelman, M., Brister, J.R. & O'Sullivan, C. (2022) The Sequence Read Archive: a decade more of explosive growth. *Nucleic Acids Research*, 50(D1), p. D387–D390. Available online: https://doi.org/10.1093/nar/gkab1053.

Keller, T.E. & Yi, S.V. (2014) DNA methylation and evolution of duplicate genes. *Proceedings of the National Academy of Sciences of the United States of America*, 111(16), p. 5932–5937. Available online: https://doi.org/10.1073/pnas.1321420111.

Khan, A., Khan, A., Ali, A., Fatima, S. & Siddiqui, M.A. (2023) Root-Knot Nematodes (Meloidogyne spp.): Biology, Plant-Nematode Interactions and Their Environmentally Benign Management Strategies. *Gesunde Pflanzen* [Preprint]. Available online: https://doi.org/10.1007/s10343-023-00886-5.

Kihika, R., Tchouassi, D.P., Ng'ang'a, M.M., Hall, D.R., Beck, J.J. & Torto, B. (2020) Compounds Associated with Infection by the Root-Knot Nematode, Meloidogyne javanica, Influence the Ability of Infective Juveniles to Recognize Host Plants. *Journal of Agricultural and Food Chemistry*, 68(34), p. 9100–9109. Available online: https://doi.org/10.1021/acs.jafc.0c03386.

Kille, B., Balaji, A., Sedlazeck, F.J., Nute, M. & Treangen, T.J. (2022) Multiple genome alignment in the telomere-to-telomere assembly era. *Genome Biology*, 23(1), p. 182. Available online: https://doi.org/10.1186/s13059-022-02735-6.

Kim, T.Y., Jang, J.Y., Yu, N.H., Chi, W.J., Bae, C.-H., Yeo, J.H., Park, A.R., Hur, J.-S., Park, H.W., Park, J.-Y., Park, J.-H., Lee, S.K. & Kim, J.-C. (2018) Nematicidal activity of grammicin produced by Xylaria grammica KCTC 13121BP against Meloidogyne incognita. *Pest Management Science*, 74(2), p. 384–391. Available online: https://doi.org/10.1002/ps.4717.

Kimura, M. (1977) Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature*, 267(5608), p. 275–276. Available online: https://doi.org/10.1038/267275a0.

Kiontke, K. & Fitch, D.H.A. (2013) Nematodes. *Current Biology: CB*, 23(19), p. R862–4. Available online: https://doi.org/10.1016/j.cub.2013.08.009.

Kiritchenko, S., Matwin, S. & Famili, F. (2005) Functional annotation of genes using hierarchical text categorization. Available online: https://research.cs.queensu.ca/biolink05/presentations/Kiritchenko.pdf.

Kirkman, L.A., Lawrence, E.A. & Deitsch, K.W. (2014) Malaria parasites utilize both homologous recombination and alternative end joining pathways to maintain genome integrity. *Nucleic Acids Research*, 42(1), p. 370–379. Available online: https://doi.org/10.1093/nar/gkt881.

Kitts, P. (2002) Genome assembly and annotation process. *McEntyre J, Ostell Jeditors. The NCBI Handbook. Bethesda: National Center for Biotechnology Information* [Preprint]. Available online: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.220.1987&rep=rep1&type=pdf.

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B.E., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J.B., Grout, J., Corlay, S. & Others (2016) Jupyter Notebooks-a publishing format for reproducible computational workflows. In *ELPUB*. books.google.com, p. 87–90. Available online: https://books.google.com/books?hl=en&lr=&id=Lgy3DAAAQBAJ&oi=fnd&pg=PA87&dq=jupyter&ots=N1ET6OqFfi&sig=HImpcIYUCZCfjy-zJlSQMBWBcv0.

Knytl, M., Fornaini, N.R., Bergelová, B., Gvoždík, V., Černohorská, H., Kubíčková, S., Fokam, E.B., Evans, B.J. & Krylov, V. (2023) Divergent subgenome evolution in the allotetraploid frog Xenopus calcaratus. *Gene*, 851, p. 146974. Available online: https://doi.org/10.1016/j.gene.2022.146974.

Kokot, M., Dlugosz, M. & Deorowicz, S. (2017) KMC 3: counting and manipulating k-mer statistics. *Bioinformatics* , 33(17), p. 2759–2761. Available online: https://doi.org/10.1093/bioinformatics/btx304.

Kolesnikova, U.K., Scott, A.D., Van de Velde, J.D., Burns, R., Tikhomirov, N.P., Pfordt, U., Clarke, A.C., Yant, L., Vekemans, X., Laurent, S. & Novikova, P.Y. (2022) Genome of selfing Siberian Arabidopsis lyrata explains establishment of allopolyploid Arabidopsis kamchatica. *bioRxiv*. Available online: https://doi.org/10.1101/2022.06.24.497443.

Kong, W., Wang, Y., Zhang, S., Yu, J. & Zhang, X. (2023) Recent advances in assembly of complex plant genomes. *Genomics, Proteomics & Bioinformatics*, 21(3), p. 427–439. Available online: https://doi.org/10.1016/j.gpb.2023.04.004.

Koren, S. & Phillippy, A.M. (2015) One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly. *Current Opinion in Microbiology*, 23, p. 110–120. Available online: https://doi.org/10.1016/j.mib.2014.11.014.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H. & Phillippy, A.M. (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research*, 27(5), p. 722–736. Available online: https://doi.org/10.1101/gr.215087.116.

Korf, I. (2004) Gene finding in novel genomes. *BMC Bioinformatics*, 5, p. 59. Available online: https://doi.org/10.1186/1471-2105-5-59.

Kosakovsky Pond, S.L. & Frost, S.D.W. (2005) Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution*, 22(5), p. 1208–1222. Available online: https://doi.org/10.1093/molbev/msi105.

Kosakovsky Pond, S.L., Poon, A.F.Y., Velazquez, R., Weaver, S., Hepler, N.L., Murrell, B., Shank, S.D., Magalis, B.R., Bouvier, D., Nekrutenko, A. & Others (2020) HyPhy 2.5—a customizable platform for evolutionary hypothesis testing using phylogenies. *Molecular Biology and Evolution*, 37(1), p. 295–299. Available online: https://academic.oup.com/mbe/article-abstract/37/1/295/5555420.

Kosiol, C., Holmes, I. & Goldman, N. (2007) An empirical codon model for protein sequence evolution. *Molecular Biology and Evolution*, 24(7), p. 1464–1479. Available online: https://doi.org/10.1093/molbev/msm064.

Köster, J. & Rahmann, S. (2012a) Building and documenting workflows with python-based snakemake. In *German Conference on Bioinformatics 2012*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik. Available online: https://drops.dagstuhl.de/opus/volltexte/2012/3717/.

Köster, J. & Rahmann, S. (2012b) Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* , 28(19), p. 2520–2522. Available online: https://doi.org/10.1093/bioinformatics/bts480.

Kostoff, D. (1939) Autosyndesis and structural hybridity in F1-hybrid Helianthus tuberosus L. x Helianthus annuus L. and their sequences. *Genetica*, 21(5-6), p. 285–300. Available online: https://doi.org/10.1007/bf01508121.

Koutsovoulos, G.D., Poullet, M., Elashry, A., Kozlowski, D.K.L., Sallet, E., Da Rocha, M., Perfus-Barbeoch, L., Martin-Jimenez, C., Frey, J.E., Ahrens, C.H., Kiewnick, S. & Danchin, E.G.J. (2020) Genome assembly and annotation of Meloidogyne enterolobii, an emerging parthenogenetic root-knot nematode. *Scientific Data*, 7(1), p. 324. Available online: https://doi.org/10.1038/s41597-020-00666-0.

Koutsovoulos, G.D., Poullet, M., El Ashry, A., Kozlowski, D.K., Sallet, E., Da Rocha, M., Martin-Jimenez, C., Perfus-Barbeoch, L., Frey, J.-E., Ahrens, C., Kiewnick, S. & Danchin, E.G.J. (2019) The polyploid genome of the mitotic parthenogenetic root-knot nematode Meloidogyne enterolobii. *bioRxiv*. Available online: https://doi.org/10.1101/586818.

Kozlowski, D.K.L., Hassanaly-Goulamhoussen, R., Da Rocha, M., Koutsovoulos, G.D., Bailly-Bechet, M. & Danchin, E.G.J. (2021) Movements of transposable elements contribute to the genomic plasticity and species diversification in an asexually reproducing nematode pest. *Evolutionary Applications*, 14(7), p. 1844–1866. Available online: https://doi.org/10.1111/eva.13246.

Kraaijeveld, A.R., Van Alphen, J.J. & Godfray, H.C. (1998) The coevolution of host resistance and parasitoid virulence. *Parasitology*, 116 Suppl, p. S29–45. Available online: https://doi.org/10.1017/s0031182000084924.

Krakauer, D.C. & Plotkin, J.B. (2002) Redundancy, antiredundancy, and the robustness of genomes. *Proceedings of the National Academy of Sciences of the United States of America*, 99(3), p. 1405–1409. Available online: https://doi.org/10.1073/pnas.032668599.

Krause, M. (1995) Transcription and translation. *Methods in Cell Biology*, 48, p. 483–512. Available online: https://www.sciencedirect.com/science/article/pii/S0091679X08614004.

Kryazhimskiy, S. & Plotkin, J.B. (2008) The population genetics of dN/dS. *PLoS Genetics*, 4(12), p. e1000304. Available online: https://doi.org/10.1371/journal.pgen.1000304.

Kryukov, K. & Imanishi, T. (2016) Human contamination in public genome assemblies. *PloS One*, 11(9), p. e0162424. Available online: https://doi.org/10.1371/journal.pone.0162424.

Kuhl, H., Du, K., Schartl, M., Kalous, L., Stöck, M. & Lamatsch, D.K. (2022) Equilibrated evolution of the mixed auto-/allopolyploid haplotype-resolved genome of the invasive hexaploid Prussian carp. *Nature Communications*, 13(1), p. 4092. Available online: https://doi.org/10.1038/s41467-022-31515-w.

Kulkarni, N., Alessandrì, L., Panero, R., Arigoni, M., Olivero, M., Ferrero, G., Cordero, F., Beccuti, M. & Calogero, R.A. (2018) Reproducible bioinformatics project: a community for reproducible bioinformatics analysis pipelines. *BMC Bioinformatics*, 19(S10). Available online: https://doi.org/10.1186/s12859-018-2296-x.

Kumar, S. & Blaxter, M.L. (2011) Simultaneous genome sequencing of symbionts and their hosts. *Symbiosis (Philadelphia, Pa.)*, 55(3), p. 119–126. Available online: https://doi.org/10.1007/s13199-012-0154-6.

Kyriakidou, M., Tai, H.H., Anglin, N.L., Ellis, D. & Strömvik, M.V. (2018) Current Strategies of Polyploid Plant Genome Sequence Assembly. *Frontiers in Plant Science*, 9, p. 1660. Available online: https://doi.org/10.3389/fpls.2018.01660.

Laetsch, D.R. & Blaxter, M.L. (2017) BlobTools: Interrogation of genome assemblies. *F1000Research*, 6(1287), p. 1287. Available online: https://doi.org/10.12688/f1000research.12232.1.

Lambshead, P.J.D. & Boucher, G. (2003) Marine nematode deep-sea biodiversity - hyperdiverse or hype?. *Journal of Biogeography*, 30(4), p. 475–485. Available online: https://doi.org/10.1046/j.1365-2699.2003.00843.x.

Lee, D.L. (2002) *The Biology of Nematodes*. CRC Press. Available online: https://play.google.com/store/books/details?id=DJu4DwAAQBAJ.

Li, A.-L., Geng, S.-F., Zhang, L.-Q., Liu, D.-C. & Mao, L. (2015) Making the Bread: Insights from Newly Synthesized Allohexaploid Wheat. *Molecular Plant*, 8(6), p. 847–859. Available online: https://doi.org/10.1016/j.molp.2015.02.016.

Li, H. (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* , 34(18), p. 3094–3100. Available online: https://doi.org/10.1093/bioinformatics/bty191.

Li, H. (2021) Concepts in phased assemblies., 17 April. Available online: https://lh3.github.io/2021/04/17/concepts-in-phased-assemblies [Accessed 08/08/2024].

Li, H. & Durbin, R. (2024) Genome assembly in the telomere-to-telomere era. *Nature Reviews. Genetics* [Preprint]. Available online: https://doi.org/10.1038/s41576-024-00718-w.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. & 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* , 25(16), p. 2078–2079. Available online: https://doi.org/10.1093/bioinformatics/btp352.

Li, M.-M., Su, Q.-L., Zu, J.-R., Xie, L., Wei, Q., Guo, H.-R., Chen, J., Zeng, R.-Z. & Zhang, Z.-S. (2022) Triploid cultivars of Cymbidium act as a bridge in the formation of polyploid plants. *Frontiers in Plant Science*, 13, p. 1029915. Available online: https://doi.org/10.3389/fpls.2022.1029915.

Lincoln, R.J., Boxshall, G.A. & Clark, P.F. (1983) A dictionary of ecology, evolution and systematics. *Systematic Botany*, 8(3), p. 339. Available online: https://doi.org/10.2307/2418488.

Li, N., Li, Y., Zheng, C., Huang, J. & Zhang, S. (2016) Genome-wide comparative analysis of the codon usage patterns in plants. *Genes & Genomics*, 38(8), p. 723–731. Available online: https://doi.org/10.1007/s13258-016-0417-3.

Liu, D., Hunt, M. & Tsai, I.J. (2018) Inferring synteny between genome assemblies: a systematic evaluation. *BMC Bioinformatics*, 19(1), p. 26. Available online: https://doi.org/10.1186/s12859-018-2026-4.

Liu, J., Hu, H., Panserat, S. & Marandel, L. (2020) Evolutionary history of DNA methylation related genes in chordates: new insights from multiple whole genome duplications. *Scientific Reports*, 10(1), p. 970. Available online: https://doi.org/10.1038/s41598-020-57753-w.

Liu, Q.L., Thomas, V.P. & Williamson, V.M. (2007) Meiotic parthenogenesis in a root-knot nematode results in rapid genomic homozygosity. *Genetics*, 176(3), p. 1483–1490. Available online: https://doi.org/10.1534/genetics.107.071134.

Li, H. (2020) *Introducing dual assembly*. Available online: https://lh3.github.io/2021/10/10/introducing-dual-assembly.

Li, Z., McKibben, M.T.W., Finch, G.S., Blischak, P.D., Sutherland, B.L. & Barker, M.S. (2021) Patterns and Processes of Diploidization in Land Plants. *Annual Review of Plant Biology*, 72, p. 387–410. Available online: https://doi.org/10.1146/annurev-arplant-050718-100344.

Logsdon, G.A., Vollger, M.R. & Eichler, E.E. (2020) Long-read human genome sequencing and its applications. *Nature Reviews. Genetics*, 21(10), p. 597–614. Available online: https://doi.org/10.1038/s41576-020-0236-x.

Loman, N.J., Quick, J. & Simpson, J.T. (2015) A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nature Methods*, 12(8), p. 733–735. Available online: https://doi.org/10.1038/nmeth.3444.

Loubalova, Z., Konstantinidou, P. & Haase, A.D. (2023) Themes and variations on piRNA-guided transposon control. *Mobile DNA*, 14(1), p. 10. Available online: https://doi.org/10.1186/s13100-023-00298-2.

Luikart, G., Kardos, M., Hand, B.K., Rajora, O.P., Aitken, S.N. & Hohenlohe, P.A. (2019) Population Genomics: Advancing Understanding of Nature. In Rajora, O.P. (ed.) *Population Genomics: Concepts, Approaches and Applications*. Cham: Springer International Publishing, p. 3–79. Available online: https://doi.org/10.1007/13836_2018_60.

Lu, J., Rincon, N., Wood, D.E., Breitwieser, F.P., Pockrandt, C., Langmead, B., Salzberg, S.L. & Steinegger, M. (2022) Metagenome analysis using the Kraken software suite. *Nature Protocols*, 17(12), p. 2815–2839. Available online: https://doi.org/10.1038/s41596-022-00738-y.

Lu, M., Li, X.-Y., Li, Z., Du, W.-X., Zhou, L., Wang, Y., Zhang, X.-J., Wang, Z.-W. & Gui, J.-F. (2021) Regain of sex determination system and sexual reproduction ability in a synthetic octoploid male fish. *Science China. Life Sciences*, 64(1), p. 77–87. Available online: https://doi.org/10.1007/s11427-020-1694-7.

Lunt, D.H. (2008) Genetic tests of ancient asexuality in root knot nematodes reveal recent hybrid origins. *BMC Evolutionary Biology*, 8, p. 194. Available online: https://doi.org/10.1186/1471-2148-8-194.

Lunt, D.H., Kumar, S., Koutsovoulos, G. & Blaxter, M.L. (2014) The complex hybrid origins of the root knot nematodes revealed through comparative genomics. *PeerJ*, 2, p. e356. Available online: https://doi.org/10.7717/peerj.356.

Luo, J., Lyu, M., Chen, R., Zhang, X., Luo, H. & Yan, C. (2019) SLR: a scaffolding algorithm based on long reads and contig classification. *BMC Bioinformatics*, 20(1), p. 1–11. Available online: https://doi.org/10.1186/s12859-019-3114-9.

Lupo, V., Van Vlierberghe, M., Vanderschuren, H., Kerff, F., Baurain, D. & Cornet, L. (2021) Contamination in reference sequence databases: Time for divide-and-rule tactics. *Frontiers in Microbiology*, 12, p. 755101. Available online: https://doi.org/10.3389/fmicb.2021.755101.

Lv, Z., Addo Nyarko, C., Ramtekey, V., Behn, H. & Mason, A.S. (2024) Defining autopolyploidy: Cytology, genetics, and taxonomy. *American Journal of Botany*, 111(8). Available online: https://doi.org/10.1002/ajb2.16292.

Lynch, M. & Conery, J.S. (2000) The evolutionary fate and consequences of duplicate genes. *Science*, 290(5494), p. 1151–1155. Available online: https://doi.org/10.1126/science.290.5494.1151.

Maccari, M., Gómez, A., Hontoria, F. & Amat, F. (2013) Functional rare males in diploid parthenogenetic Artemia. *Journal of Evolutionary Biology*, 26(9), p. 1934–1948. Available online: https://doi.org/10.1111/jeb.12191.

Maciver, S.K. (2016) Asexual Amoebae Escape Muller's Ratchet through Polyploidy. *Trends in Parasitology*, 32(11), p. 855–862. Available online: https://doi.org/10.1016/j.pt.2016.08.006.

Mandáková, T., Gloss, A.D., Whiteman, N.K. & Lysak, M.A. (2016) How diploidization turned a tetraploid into a pseudotriploid. *American Journal of Botany*, 103(7), p. 1187–1196. Available online: https://doi.org/10.3732/ajb.1500452.

Mandal, H.R., Katel, S., Subedi, S. & Shrestha, J. (2021) Plant Parasitic Nematodes and their management in crop production: a review. *Journal of Agriculture and Natural Resources*, 4(2), p. 327–338. Available online: https://www.academia.edu/download/74216329/26664.pdf.

Mansai, S.P. & Innan, H. (2010) The power of the methods for detecting interlocus gene conversion. *Genetics*, 184(2), p. 517–527. Available online: https://doi.org/10.1534/genetics.109.111161.

Mapleson, D., Garcia, A.G., Kettleborough, G., Wright, J. & Clavijo, B.J. (2017) KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics*, 33(4), p. 574–576. Available online: https://academic.oup.com/bioinformatics/article-abstract/33/4/574/2664339.

Marçais, G., Delcher, A.L., Phillippy, A.M., Coston, R., Salzberg, S.L. & Zimin, A. (2018) MUMmer4: A fast and versatile genome alignment system. *PLoS Computational Biology*, 14(1), p. e1005944. Available online: https://doi.org/10.1371/journal.pcbi.1005944.

Marçais, G. & Kingsford, C. (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* , 27(6), p. 764–770. Available online: https://doi.org/10.1093/bioinformatics/btr011.

Marquez, J. & Hajihassani, A. (2023) Successional effects of cover cropping and deep tillage on suppression of plant-parasitic nematodes and soilborne fungal pathogens. *Pest Management Science*, 79(8), p. 2737–2747. Available online: https://doi.org/10.1002/ps.7450.

Martin, E., Kahraman, A., Dirmenci, T., Bozkurt, H. & Eroğlu, H. (2020) Karyotype evolution and new chromosomal data in Erodium: chromosome alteration, polyploidy, dysploidy, and symmetrical karyotypes. *Turkish Journal of Botany* [Preprint]. Available online: https://doi.org/10.3906/bot-1912-22.

Mason, A.S. & Chris Pires, J. (2015) Unreduced gametes: meiotic mishap or evolutionary mechanism?. *Trends in Genetics: TIG*, 31(1), p. 5–10. Available online: https://doi.org/10.1016/j.tig.2014.09.011.

Mason, A.S. & Wendel, J.F. (2020) Homoeologous Exchanges, Segmental Allopolyploidy, and Polyploid Genome Evolution. *Frontiers in Genetics*, 11, p. 1014. Available online: https://doi.org/10.3389/fgene.2020.01014.

Mason, J.M., Randall, T.A. & Capkova Frydrychova, R. (2016) Telomerase lost?. *Chromosoma*, 125(1), p. 65–73. Available online: https://doi.org/10.1007/s00412-015-0528-7.

Mason, J.M., Reddy, H.M. & Frydrychova, R.C. (2011) Telomere Maintenance in Organisms without Telomerase. In Seligmann, H. (ed.) *DNA Replication*. Rijeka: IntechOpen. Available online: https://doi.org/10.5772/19348.

Masterson, J. (1994) Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. *Science*, 264(5157), p. 421–424. Available online: https://doi.org/10.1126/science.264.5157.421.

Mathers, T.C., Wouters, R.H.M., Mugford, S.T., Swarbreck, D., van Oosterhout, C. & Hogenhout, S.A. (2021) Chromosome-scale genome assemblies of aphids reveal extensively rearranged autosomes and long-term conservation of the X chromosome. *Molecular Biology and Evolution*, 38(3), p. 856–875. Available online: https://doi.org/10.1093/molbev/msaa246.

Ma, X.-F. & Gustafson, J.P. (2005) Genome evolution of allopolyploids: a process of cytological and genetic diploidization. *Cytogenetic and Genome Research*, 109(1-3), p. 236–249. Available online: https://doi.org/10.1159/000082406.

McClintock, B. (1984) The significance of responses of the genome to challenge. *Science (New York, N.Y.)*, 226(4676), p. 792–801. Available online: https://doi.org/10.1126/science.15739260.

McGuire, A.L., Gabriel, S., Tishkoff, S.A., Wonkam, A., Chakravarti, A., Furlong, E.E.M., Treutlein, B., Meissner, A., Chang, H.Y., López-Bigas, N., Segal, E. & Kim, J.-S. (2020) The road ahead in genetics and genomics. *Nature Reviews. Genetics*, 21(10), p. 581–596. Available online: https://doi.org/10.1038/s41576-020-0272-6.

Meirmans, P.G., Liu, S. & van Tienderen, P.H. (2018) The Analysis of Polyploid Genetic Data. *The Journal of Heredity*, 109(3), p. 283–296. Available online: https://doi.org/10.1093/jhered/esy006.

Mejias, J., Truong, N.M., Abad, P., Favery, B. & Quentin, M. (2019) Plant Proteins and Processes Targeted by Parasitic Nematode Effectors. *Frontiers in Plant Science*, 10, p. 970. Available online: https://doi.org/10.3389/fpls.2019.00970.

Mendes, F.K. & Hahn, M.W. (2016) Gene Tree Discordance Causes Apparent Substitution Rate Variation. *Systematic Biology*, 65(4), p. 711–721. Available online: https://doi.org/10.1093/sysbio/syw018.

Meng, Y., Lei, Y., Gao, J., Liu, Y., Ma, E., Ding, Y., Bian, Y., Zu, H., Dong, Y. & Zhu, X. (2022) Genome sequence assembly algorithms and misassembly identification methods. *Molecular Biology Reports*, 49(11), p. 11133–11148. Available online: https://doi.org/10.1007/s11033-022-07919-8.

Merchant, S., Wood, D.E. & Salzberg, S.L. (2014) Unexpected cross-species contamination in genome sequencing projects. *PeerJ*, 2, p. e675. Available online: https://doi.org/10.7717/peerj.675.

Miga, K.H. & Eichler, E.E. (2023) Envisioning a new era: Complete genetic information from routine, telomere-to-telomere genomes. *The American Journal of Human Genetics*, 110(11), p. 1832–1840. Available online: https://doi.org/10.1016/j.ajhg.2023.09.011.

Milligan, S.B., Bodeau, J., Yaghoobi, J., Kaloshian, I., Zabel, P. & Williamson, V.M. (1998) The root knot nematode resistance gene Mi from tomato is a member of the leucine zipper, nucleotide binding, leucine-rich repeat family of plant genes. *The Plant Cell*, 10(8), p. 1307–1319. Available online: https://doi.org/10.1105/tpc.10.8.1307.

Ming, R. & Man Wai, C. (2015) Assembling allopolyploid genomes: no longer formidable. *Genome Biology*, 16, p. 27. Available online: https://doi.org/10.1186/s13059-015-0585-5.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A. & Lanfear, R. (2020) IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, 37(5), p. 1530–1534. Available online: https://doi.org/10.1093/molbev/msaa015.

Misas-Villamil, J.C., van der Hoorn, R.A.L. & Doehlemann, G. (2016) Papain-like cysteine proteases as hubs in plant immunity. *The New Phytologist*, 212(4), p. 902–907. Available online: https://doi.org/10.1111/nph.14117.

Mitiku, M. (2018) Plant-parasitic nematodes and their management: A review. *Agric. Res. Technol* [Preprint]. Available online: https://pdfs.semanticscholar.org/e184/cbf7b64eb25bf08bfb6ecc94a45d9931d672.pdf.

Mitkowski, N.A. & Abawi, G.S. (2003) Root-knot nematode. *The American Phytopathological Scoiety* [Preprint]. Available online: https://doi.org/10.1094/PHI-I-2003-0917-01.

Mitros, T., Lyons, J.B., Session, A.M., Jenkins, J., Shu, S., Kwon, T., Lane, M., Ng, C., Grammer, T.C., Khokha, M.K., Grimwood, J., Schmutz, J., Harland, R.M. & Rokhsar, D.S. (2019) A chromosome-scale genome assembly and dense genetic map for Xenopus tropicalis. *Developmental Biology*, 452(1), p. 8–20. Available online: https://doi.org/10.1016/j.ydbio.2019.03.015.

Mogie, M. (1986) Automixis: its distribution and status. *Biological Journal of the Linnean Society. Linnean Society of London*, 28, p. 321–329. Available online: https://doi.org/10.1111/J.1095-8312.1986.TB01761.X.

Moore, G. (2002) Meiosis in allopolyploids – the importance of 'Teflon' chromosomes. *Trends in Genetics: TIG*, 18(9), p. 456–463. Available online: https://doi.org/10.1016/S0168-9525(02)02730-0.

Moosavi, M.R. (2015) Damage of the root-knot nematode Meloidogyne javanica to bell pepper, Capsicum annuum. *Journal of Plant Diseases and Protection: Scientific Journal of the German Phytomedical Society* , 122(5/6), p. 244–249. Available online: http://www.jstor.org/stable/24618902.

Mortier, F., Bafort, Q., Milosavljevic, S., Kauai, F., Prost Boxoen, L., Van de Peer, Y. & Bonte, D. (2024) Understanding polyploid establishment: temporary persistence or stable coexistence?. *Oikos* , 2024(5). Available online: https://doi.org/10.1111/oik.09929.

Mota, A.P.Z., Koutsovoulos, G.D., Perfus-Barbeoch, L., Despot-Slade, E., Labadie, K., Aury, J.-M., Robbe-Sermesant, K., Bailly-Bechet, M., Belser, C., Péré, A., Rancurel, C., Kozlowski, D.K., Hassanaly-Goulamhoussen, R., Da Rocha, M., Noel, B., Meštrović, N., Wincker, P. & Danchin, E.G.J. (2024) Unzipped genome assemblies of polyploid root-knot nematodes reveal unusual and clade-specific telomeric repeats. *Nature Communications*, 15(1), p. 773. Available online: https://doi.org/10.1038/s41467-024-44914-y.

Moulé, L. (1911) La parasitologie dans la littérature antique. II. Les parasites du tube digestif. *Archives de Parasitologie* [Preprint]. Available online: https://scholar.google.ca/scholar?cluster=9132594585644517768&hl=en&as_sdt=0,5&sciodt=0,5.

Muller, H.J. (1932) Some Genetic Aspects of Sex. *The American Naturalist*, 66(703), p. 118–138. Available online: https://doi.org/10.1086/280418.

Murrell, B., Weaver, S., Smith, M.D., Wertheim, J.O., Murrell, S., Aylward, A., Eren, K., Pollner, T., Martin, D.P., Smith, D.M., Scheffler, K. & Kosakovsky Pond, S.L. (2015) Gene-wide identification of episodic selection. *Molecular Biology and Evolution*, 32(5), p. 1365–1371. Available online: https://doi.org/10.1093/molbev/msv035.

Murrell, B., Wertheim, J.O., Moola, S., Weighill, T., Scheffler, K. & Kosakovsky Pond, S.L. (2012) Detecting individual sites subject to episodic diversifying selection. *PLoS Genetics*, 8(7), p. e1002764. Available online: https://doi.org/10.1371/journal.pgen.1002764.

Muse, S.V. & Gaut, B.S. (1994) A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Molecular Biology and Evolution*, 11(5), p. 715–724. Available online: https://doi.org/10.1093/oxfordjournals.molbev.a040152.

Narayanasamy, P. (2002) *Microbial Plant Pathogens and Crop Disease Management*. https://doi.org/10.1201/9781482279948: CRC Press.

Navarrete, M., Djian-Caporalino, C., Mateille, T., Palloix, A., Sage-Palloix, A.-M., Lefèvre, A., Fazari, A., Marteu, N., Tavoillot, J., Dufils, A., Furnion, C., Pares, L. & Forest, I. (2016) A resistant pepper used as a trap cover crop in vegetable production strongly decreases root-knot nematode infestation in soil. *Agronomy for Sustainable Development*, 36(4). Available online: https://doi.org/10.1007/s13593-016-0401-y.

Neiman, M., Sharbel, T.F. & Schwander, T. (2014) Genetic causes of transitions from sexual reproduction to asexuality in plants and animals. *Journal of Evolutionary Biology*, 27(7), p. 1346–1359. Available online: https://doi.org/10.1111/jeb.12357.

Nguyen, C.-N., Perfus-Barbeoch, L., Quentin, M., Zhao, J., Magliano, M., Marteu, N., Da Rocha, M., Nottet, N., Abad, P. & Favery, B. (2018) A root-knot nematode small glycine and cysteine-rich secreted effector, MiSGCR1, is involved in plant parasitism. *The New Phytologist*, 217(2), p. 687–699. Available online: https://doi.org/10.1111/nph.14837.

Noack, S., Harrington, J., Carithers, D.S., Kaminsky, R. & Selzer, P.M. (2021) Heartworm disease - Overview, intervention, and industry perspective. *International Journal for Parasitology, Drugs and Drug Resistance*, 16, p. 65–89. Available online: https://doi.org/10.1016/j.ijpddr.2021.03.004.

Noling, J. (2014) Nematode management in tomatoes, peppers, and eggplant. Available online: https://www.semanticscholar.org/paper/Nematode-management-in-tomatoes%2C-peppers%2C-and-Noling/7cc14631b29dac4c771e209faab5d2eec3f5f98a [Accessed 06/08/2024].

Noling, J.W. (2000) Effects of continuous culture of a resistant tomato cultivar on Meloidogyne incognita soil population density and pathogenicity. *Journal of Nematology* [Preprint].

Nombela, G., Williamson, V.M. & Muñiz, M. (2003) The root-knot nematode resistance gene Mi-1.2 of tomato is responsible for resistance against the whitefly Bemisia tabaci. *Molecular Plant-Microbe Interactions: MPMI*, 16(7), p. 645–649. Available online: https://doi.org/10.1094/MPMI.2003.16.7.645.

Nouri, N.V., Rahmatian, R. & Salehi, A. (2022) Prevalence of Helminthic Infections in the Gastrointestinal Tract of Cattle in Mazandaran Province (Northern Iran). *Journal of Parasitology Research*, 2022, p. 7424647. Available online: https://doi.org/10.1155/2022/7424647.

O'Doherty, K.C., Shabani, M., Dove, E.S., Bentzen, H.B., Borry, P., Burgess, M.M., Chalmers, D., De Vries, J., Eckstein, L., Fullerton, S.M., Juengst, E., Kato, K., Kaye, J., Knoppers, B.M., Koenig, B.A., Manson, S.M., McGrail, K.M., McGuire, A.L., Meslin, E.M., Nicol, D., Prainsack, B., Terry, S.F., Thorogood, A. & Burke, W. (2021) Toward better governance of human genomic data. *Nature Genetics*, 53(1), p. 2–8. Available online: https://doi.org/10.1038/s41588-020-00742-6.

Oka, Y., Ben-Daniel, B. & Cohen, Y. (2012) Nematicidal activity of the leaf powder and extracts ofMyrtus communisagainst the root-knot nematodeMeloidogyne javanica. *Plant Pathology*, 61(6), p. 1012–1020. Available online: https://doi.org/10.1111/j.1365-3059.2011.02587.x.

Oleszczuk, S. & Lukaszewski, A.J. (2014) The origin of unusual chromosome constitutions among newly formed allopolyploids. *American Journal of Botany* [Preprint]. Available online: https://bsapubs.onlinelibrary.wiley.com/doi/abs/10.3732/ajb.1300286.

Olivier, M., Gregory, D.J. & Forget, G. (2005) Subversion mechanisms by which Leishmania parasites can escape the host immune response: a signaling point of view. *Clinical Microbiology Reviews*, 18(2), p. 293–305. Available online: https://doi.org/10.1128/CMR.18.2.293-305.2005.

Olsen, M.W. (2000) Root knot nematode. Available online: https://cales.arizona.edu/forageandgrain/sites/cals.arizona.edu.forageandgrain/files/az1187.pdf.

Ondov, B.D., Treangen, T.J., Melsted, P., Mallonee, A.B., Bergman, N.H., Koren, S. & Phillippy, A.M. (2016) Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biology*, 17(1), p. 132. Available online: https://doi.org/10.1186/s13059-016-0997-x.

Ornat, C., Verdejo-Lucas, S. & Sorribas, F.J. (2001) A population of Meloidogyne javanica in Spain virulent to the Mi resistance gene in tomato. *Plant Disease*, 85(3), p. 271–276. Available online: https://doi.org/10.1094/PDIS.2001.85.3.271.

Osborn, T.C., Pires, J.C., Birchler, J.A., Auger, D.L., Chen, Z.J., Lee, H.-S., Comai, L., Madlung, A., Doerge, R.W., Colot, V. & Martienssen, R.A. (2003) Understanding mechanisms of novel gene expression in polyploids. *Trends in Genetics: TIG*, 19(3), p. 141–147. Available online: https://doi.org/10.1016/s0168-9525(03)00015-5.

Otto, S.P. & Whitton, J. (2000) Polyploid incidence and evolution. *Annual Review of Genetics*, 34, p. 401–437. Available online: https://doi.org/10.1146/annurev.genet.34.1.401.

PacificBiosciences (2022) *IsoSeq3*. Github. Available online: https://github.com/PacificBiosciences/IsoSeq [Accessed 21/10/2022].

Papadopoulou, J. & Traintaphyllou, A.C. (1982) Sex differentiation in Meloidogyne incognita and anatomical evidence of sex reversal. *Journal of Nematology*, 14(4), p. 549–566. Available online: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2618216/.

Paquin, C. & Adams, J. (1983) Frequency of fixation of adaptive mutations is higher in evolving diploid than haploid yeast populations. *Nature*, 302(5908), p. 495–500. Available online: https://doi.org/10.1038/302495a0.

Paradis, E. (2010) Pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics* , 26(3), p. 419–420. Available online: https://doi.org/10.1093/bioinformatics/btp696.

Pardue, M.-L. & DeBaryshe, P.G. (2011) Retrotransposons that maintain chromosome ends. *Proceedings of the National Academy of Sciences of the United States of America*, 108(51), p. 20317–20324. Available online: https://doi.org/10.1073/pnas.1100278108.

Parra, G., Bradnam, K. & Korf, I. (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* , 23(9), p. 1061–1067. Available online: https://doi.org/10.1093/bioinformatics/btm071.

Parvathy, S.T., Udayasuriyan, V. & Bhadana, V. (2022) Codon usage bias. *Molecular Biology Reports*, 49(1), p. 539–565. Available online: https://doi.org/10.1007/s11033-021-06749-4.

Pascual, S., Emiliozzi, M. & Nombela, G. (2024) Role of Two Transcription Factors (TGA 1a and TGA 2.1) in the Mi-1-Mediated Resistance of Tomato to the Root-Knot Nematode Meloidogyne javanica. *Horticulturae*, 10(2), p. 134. Available online: https://doi.org/10.3390/horticulturae10020134.

Pavlidis, P. & Alachiotis, N. (2017) A survey of methods and tools to detect recent and strong positive selection. *Journal of Biological Research (Thessalonike, Greece)*, 24(1), p. 7. Available online: https://doi.org/10.1186/s40709-017-0064-0.

Pearson, W.R. & Lipman, D.J. (1988) Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences of the United States of America*, 85(8), p. 2444–2448. Available online: https://doi.org/10.1073/pnas.85.8.2444.

Peden, J. (2005) CodonW version 1.4. 2. *University of Nottingham: Nottingham, UK* [Preprint].

Peden, J.F. (2000) *Analysis of codon usage*. Doctor of Philosophy. University of Nottingham. Available online: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=450270ddf16353d87927 4211d08b9fa7f1ea5537.

Pellino, M., Hojsgaard, D., Schmutzer, T., Scholz, U., Hörandl, E., Vogel, H. & Sharbel, T.F. (2013) Asexual genome evolution in the apomictic Ranunculus auricomus complex: examining the effects of hybridization and mutation accumulation. *Molecular Ecology*, 22(23), p. 5908–5921. Available online: https://doi.org/10.1111/mec.12533.

Peng, R.D. & Hicks, S.C. (2021) Reproducible research: A retrospective. *Annual Review of Public Health*, 42(1), p. 79–93. Available online: https://doi.org/10.1146/annurev-publhealth-012420-105110.

Peona, V., Blom, M.P.K., Xu, L., Burri, R., Sullivan, S., Bunikis, I., Liachko, I., Haryoko, T., Jønsson, K.A., Zhou, Q., Irestedt, M. & Suh, A. (2021) Identifying the causes and consequences of assembly gaps using a multiplatform genome assembly of a bird-of-paradise. *Molecular Ecology Resources*, 21(1), p. 263–286. Available online: https://doi.org/10.1111/1755-0998.13252.

Perfus-Barbeoch, L., Castagnone-Sereno, P., Reichelt, M., Fneich, S., Roquis, D., Pratx, L., Cosseau, C., Grunau, C. & Abad, P. (2014) Elucidating the molecular bases of epigenetic inheritance in non-model invertebrates: the case of the root-knot nematode Meloidogyne incognita. *Frontiers in Physiology*, 5, p. 211. Available online: https://doi.org/10.3389/fphys.2014.00211.

Perry, R.N., Moens, M. & Starr, J.L. (2009) *Root-knot Nematodes*. CABI. Available online: https://play.google.com/store/books/details?id=UN3uHMr_UCoC.

Pflüger, M. (2019) Reproducible Data Analysis with conda. In *Open Science Days 2019*. pure.mpg.de. Available online: https://pure.mpg.de/rest/items/item_3027294/component/file_3027295/content.

Ploeg, A.T., Stoddard, C.S., Turini, T.A., Nunez, J.J., Miyao, E.M. & Subbotin, S.A. (2023) Tomato Mi-gene Resistance-Breaking Populations of Meloidogyne Show Variable Reproduction on Susceptible and Resistant Crop Cultivars. *Journal of Nematology*, 55(1), p. 20230043. Available online: https://doi.org/10.2478/jofnem-2023-0043.

Pogorelko, G.V., Juvale, P.S., Rutter, W.B., Hütten, M., Maier, T.R., Hewezi, T., Paulus, J., van der Hoorn, R.A., Grundler, F.M., Siddique, S., Lionetti, V., Zabotina, O.A. & Baum, T.J. (2019) Re-targeting of a plant defense protease by a cyst nematode effector. *The Plant Journal: For Cell and Molecular Biology*, 98(6), p. 1000–1014. Available online: https://doi.org/10.1111/tpj.14295.

Pop, M. (2009) Genome assembly reborn: recent computational challenges. *Briefings in Bioinformatics*, 10(4), p. 354–366. Available online: https://doi.org/10.1093/bib/bbp026.

Porazinska, D.L., Giblin-Davis, R.M., Powers, T.O. & Thomas, W.K. (2012) Nematode spatial and ecological patterns from tropical and temperate rainforests. *PloS One*, 7(9), p. e44641. Available online: https://doi.org/10.1371/journal.pone.0044641.

Poulin, R. (2011) The many roads to parasitism: a tale of convergence. *Advances in Parasitology*, 74, p. 1–40. Available online: https://doi.org/10.1016/B978-0-12-385897-9.00001-X.

Poulin, R. & Randhawa, H.S. (2015) Evolution of parasitism along convergent lines: from ecology to genomics. *Parasitology*, 142 Suppl 1(S1), p. S6–S15. Available online: https://doi.org/10.1017/S0031182013001674.

Pratx, L., Rancurel, C., Da Rocha, M., Danchin, E.G.J., Castagnone-Sereno, P., Abad, P. & Perfus-Barbeoch, L. (2018) Genome-wide expert annotation of the epigenetic machinery of the plant-parasitic nematodes Meloidogyne spp., with a focus on the asexually reproducing species. *BMC Genomics*, 19(1), p. 321. Available online: https://doi.org/10.1186/s12864-018-4686-x.

Qian, L., Qian, W. & Snowdon, R.J. (2014) Sub-genomic selection patterns as a signature of breeding in the allopolyploid Brassica napus genome. *BMC Genomics*, 15(1), p. 1170. Available online: https://doi.org/10.1186/1471-2164-15-1170.

Rabanus-Wallace, M.T., Hackauf, B., Mascher, M., Lux, T., Wicker, T., Gundlach, H., Baez, M., Houben, A., Mayer, K.F.X., Guo, L., Poland, J., Pozniak, C.J., Walkowiak, S., Melonek, J., Praz, C.R., Schreiber, M., Budak, H., Heuberger, M., Steuernagel, B., Wulff, B., Börner, A., Byrns, B., Čížková, J., Fowler, D.B., Fritz, A., Himmelbach, A., Kaithakottil, G., Keilwagen, J., Keller, B., Konkin, D., Larsen, J., Li, Q., Myśków, B., Padmarasu, S., Rawat, N., Sesiz, U., Biyiklioglu-Kaya, S., Sharpe, A., Šimková, H., Small, I., Swarbreck, D., Toegelová, H., Tsvetkova, N., Voylokov, A.V., Vrána, J., Bauer, E., Bolibok-Bragoszewska, H., Doležel, J., Hall, A., Jia, J., Korzun, V., Laroche, A., Ma, X.-F., Ordon, F., Özkan, H., Rakoczy-Trojanowska, M., Scholz, U., Schulman, A.H., Siekmann, D., Stojałowski, S., Tiwari, V.K., Spannagl, M. & Stein, N. (2021) Chromosome-scale genome assembly provides insights into rye biology, evolution and agronomic potential. *Nature Genetics*, 53(4), p. 564–573. Available online: https://doi.org/10.1038/s41588-021-00807-0.

Ralmi, N.H.A.A., Khandaker, M.M., Mat, N. & Others (2016) Occurrence and control of root knot nematode in crops: a review. *Australian Journal of Crop Science*, 11(12), p. 1649. Available online: https://search.informit.com.au/documentSummary;dn=581347147481437;res=IELHSS.

Ralmi, N.H.A.A., School of Agriculture Science and Biotechnology, Faculty of Bioresources and Food Industry, University of Sultan Zainal Abidin, Besut Campus, 22000 Besut, Terengganu, Malaysia, Khandaker, M.M., Mat, N., School of Agriculture Science and Biotechnology, Faculty of Bioresources and Food Industry, University of Sultan Zainal Abidin, Besut Campus, 22000 Besut, Terengganu, Malaysia & School of Agriculture Science and Biotechnology, Faculty of Bioresources and Food Industry, University of Sultan Zainal Abidin, Besut Campus, 22000 Besut, Terengganu, Malaysia (2016) Occurrence and control of root knot nematode in crops: A review. *Australian Journal of Crop Science*, 10(12), p. 1649–1654. Available online: https://doi.org/10.21475/ajcs.2016.10.12.p7444.

Ram, K. (2013) Git can facilitate greater reproducibility and increased transparency in science. *Source Code for Biology and Medicine*, 8(1), p. 7. Available online: https://doi.org/10.1186/1751-0473-8-7.

Rammah, A. & Hirschmann, H. (1990) Morphological Comparison of Three Host Races of Meloidogyne javanica. *Journal of Nematology*, 22(1), p. 56–68. Available online: https://www.ncbi.nlm.nih.gov/pubmed/19287689.

Ramsey, J. & Schemske, D.W. (1998) PATHWAYS, MECHANISMS, AND RATES OF POLYPLOID FORMATION IN FLOWERING PLANTS. *Annual Review of Ecology, Evolution, and Systematics*, 29(Volume 29, 1998), p. 467–501. Available online: https://doi.org/10.1146/annurev.ecolsys.29.1.467.

Ranallo-Benavidez, T.R., Jaron, K.S. & Schatz, M.C. (2020) GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications*, 11(1), p. 1432. Available online: https://doi.org/10.1038/s41467-020-14998-3.

Reddick, L.E. & Alto, N.M. (2014) Bacteria fighting back: how pathogens target and subvert the host innate immune system. *Molecular Cell*, 54(2), p. 321–328. Available online: https://doi.org/10.1016/j.molcel.2014.03.010.

Reid, K., Bell, M.A. & Veeramah, K.R. (2021) Threespine Stickleback: A Model System For Evolutionary Genomics. *Annual Review of Genomics and Human Genetics*, 22(1), p. 357–383. Available online: https://doi.org/10.1146/annurev-genom-111720-081402.

Reilly, J., Tubiello, F., McCarl, B. & Melillo, J. (2001) Climate change and agriculture in the United States. *Climate Change Impacts on the United States: US National Assessment of the Potential Consequences of Climate Variability and Change: Foundation*, p. 379–403. Available online: https://www.academia.edu/download/76792679/13Agri.pdf.

Rhie, A., McCarthy, S.A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., Uliano-Silva, M., Chow, W., Fungtammasan, A., Kim, J., Lee, C., Ko, B.J., Chaisson, M., Gedman, G.L., Cantin, L.J., Thibaud-Nissen, F., Haggerty, L., Bista, I., Smith, M., Haase, B., Mountcastle, J., Winkler, S., Paez, S., Howard, J., Vernes, S.C., Lama, T.M., Grutzner, F., Warren, W.C., Balakrishnan, C.N., Burt, D., George, J.M., Biegler, M.T., Iorns, D., Digby, A., Eason, D., Robertson, B., Edwards, T., Wilkinson, M., Turner, G., Meyer, A., Kautt, A.F., Franchini, P., Detrich, H.W., Svardal, H., Wagner, M., Naylor, G.J.P., Pippel, M., Malinsky, M., Mooney, M., Simbirsky, M., Hannigan, B.T., Pesout, T., Houck, M., Misuraca, A., Kingan, S.B., Hall, R., Kronenberg, Z., Sović, I., Dunn, C., Ning, Z., Hastie, A., Lee, J., Selvaraj, S., Green, R.E., Putnam, N.H., Gut, I., Ghurye, J., Garrison, E., Sims, Y., Collins, J., Pelan, S., Torrance, J., Tracey, A., Wood, J., Dagnew, R.E., Guan, D., London, S.E., Clayton, D.F., Mello, C.V., Friedrich, S.R., Lovell, P.V., Osipova, E., Al-Ajli, F.O., Secomandi, S., Kim, H., Theofanopoulou, C., Hiller, M., Zhou, Y., Harris, R.S., Makova, K.D., Medvedev, P., Hoffman, J., Masterson, P., Clark, K., Martin, F., Howe, K., Flicek, P., Walenz, B.P., Kwak, W., Clawson, H., Diekhans, M., Nassar, L., Paten, B., Kraus, R.H.S., Crawford, A.J., Gilbert, M.T.P., Zhang, G., Venkatesh, B., Murphy, R.W., Koepfli, K.-P., Shapiro, B., Johnson, W.E., Di Palma, F., Marques-Bonet, T., Teeling, E.C., Warnow, T., Graves, J.M., Ryder, O.A., Haussler, D., O'Brien, S.J., Korlach, J., Lewin, H.A., Howe, K., Myers, E.W., Durbin, R., Phillippy, A.M. & Jarvis, E.D. (2021) Towards complete and error-free genome assemblies of all vertebrate species. *Nature*, 592(7856), p. 737–746. Available online: https://doi.org/10.1038/s41586-021-03451-0.

Rhie, A., Walenz, B.P., Koren, S. & Phillippy, A.M. (2020) Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology*, 21(1), p. 245. Available online: https://doi.org/10.1186/s13059-020-02134-9.

Rhoads, A. & Au, K.F. (2015) PacBio Sequencing and Its Applications. *Genomics, Proteomics & Bioinformatics*, 13(5), p. 278–289. Available online: https://doi.org/10.1016/j.gpb.2015.08.002.

Rice, E.S. & Green, R.E. (2019) New approaches for genome assembly and scaffolding. *Annual Review of Animal Biosciences*, 7(1), p. 17–40. Available online: https://doi.org/10.1146/annurev-animal-020518-115344.

Roberts, P.A. & Thomason, J. (1986) *Variability in reproduction of isolates of Meloidogyne incognita and M. javanica on resistant tomato genotypes*. worldveg.tind.io. Available online: https://worldveg.tind.io/record/4772/.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. & Mesirov, J.P. (2011) Integrative genomics viewer. *Nature Biotechnology*, 29(1), p. 24–26. Available online: https://doi.org/10.1038/nbt.1754.

Rokas, A. & Abbot, P. (2009) Harnessing genomics for evolutionary insights. *Trends in Ecology & Evolution*, 24(4), p. 192–200. Available online: https://doi.org/10.1016/j.tree.2008.11.004.

Rudolphi, K.A. (2008) Neue Beobachtungen über die Eingeweidewürmer. *Archiv Für Zoologie Und Zootomie*, 3, p. 1– 32.

Rudolph, R.E., Bajek, V. & Munir, M. (2023) Effects of soil solarization and grafting on tomato yield and southern root-knot nematode population densities. *HortScience: A Publication of the American Society for Horticultural Science*, 58(11), p. 1443–1449. Available online: https://doi.org/10.21273/hortsci17396-23.

Rutter, W.B., Franco, J. & Gleason, C. (2022) Rooting out the mechanisms of root-knot nematode-plant interactions. *Annual Review of Phytopathology*, 60(1), p. 43–76. Available online: https://doi.org/10.1146/annurev-phyto-021621-120943.

Saada, O.A., Friedrich, A. & Schacherer, J. (2022) Towards accurate, contiguous and complete alignment-based polyploid phasing algorithms. *Genomics*, 114(3), p. 110369. Available online: https://doi.org/10.1016/j.ygeno.2022.110369.

Salojärvi, J., Rambani, A., Yu, Z., Guyot, R., Strickler, S., Lepelley, M., Wang, C., Rajaraman, S., Rastas, P., Zheng, C., Muñoz, D.S., Meidanis, J., Paschoal, A.R., Bawin, Y., Krabbenhoft, T.J., Wang, Z.Q., Fleck, S.J., Aussel, R., Bellanger, L., Charpagne, A., Fournier, C., Kassam, M., Lefebvre, G., Métairon, S., Moine, D., Rigoreau, M., Stolte, J., Hamon, P., Couturon, E., Tranchant-Dubreuil, C., Mukherjee, M., Lan, T., Engelhardt, J., Stadler, P., Correia De Lemos, S.M., Suzuki, S.I., Sumirat, U., Wai, C.M., Dauchot, N., Orozco-Arias, S., Garavito, A., Kiwuka, C., Musoli, P., Nalukenge, A., Guichoux, E., Reinout, H., Smit, M., Carretero-Paulet, L., Filho, O.G., Braghini, M.T., Padilha, L., Sera, G.H., Ruttink, T., Henry, R., Marraccini, P., Van de Peer, Y., Andrade, A., Domingues, D., Giuliano, G., Mueller, L., Pereira, L.F., Plaisance, S., Poncet, V., Rombauts, S., Sankoff, D., Albert, V.A., Crouzillat, D., de Kochko, A. & Descombes, P. (2024) The genome and population genomics of allopolyploid Coffea arabica reveal the diversification history of modern coffee cultivars. *Nature Genetics*, 56(4), p. 721–731. Available online: https://doi.org/10.1038/s41588-024-01695-w.

Salter, S.J., Cox, M.J., Turek, E.M., Calus, S.T., Cookson, W.O., Moffatt, M.F., Turner, P., Parkhill, J., Loman, N.J. & Walker, A.W. (2014) Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology*, 12, p. 87. Available online: https://doi.org/10.1186/s12915-014-0087-z.

Sasaki, T. & Fujiwara, H. (2000) Detection and distribution patterns of telomerase activity in insects: Telomerase activity in insects. *European Journal of Biochemistry*, 267(10), p. 3025–3031. Available online: https://doi.org/10.1046/j.1432-1033.2000.01323.x.

Sati, S. & Cavalli, G. (2017) Chromosome conformation capture technologies and their impact in understanding genome function. *Chromosoma*, 126(1), p. 33–44. Available online: https://doi.org/10.1007/s00412-016-0593-6.

Sato, K., Kadota, Y., Gan, P., Bino, T., Uehara, T., Yamaguchi, K., Ichihashi, Y., Maki, N., Iwahori, H., Suzuki, T., Shigenobu, S. & Shirasu, K. (2018) High-Quality Genome Sequence of the Root-Knot Nematode Meloidogyne arenaria Genotype A2-O. *Genome Announcements*, 6(26). Available online: https://doi.org/10.1128/genomeA.00519-18.

Sattler, M.C., Carvalho, C.R. & Clarindo, W.R. (2016) The polyploidy and its key role in plant breeding. *Planta*, 243(2), p. 281–296. Available online: https://doi.org/10.1007/s00425-015-2450-x.

Sayers, E.W., Bolton, E.E., Brister, J.R., Canese, K., Chan, J., Comeau, D.C., Connor, R., Funk, K., Kelly, C., Kim, S., Madej, T., Marchler-Bauer, A., Lanczycki, C., Lathrop, S., Lu, Z., Thibaud-Nissen, F., Murphy, T., Phan, L., Skripchenko, Y., Tse, T., Wang, J., Williams, R., Trawick, B.W., Pruitt, K.D. & Sherry, S.T. (2022) Database resources of the national center for biotechnology information. *Nucleic Acids Research*, 50(D1), p. D20–D26. Available online: https://doi.org/10.1093/nar/gkab1112.

Sayers, E.W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K.D. & Karsch-Mizrachi, I. (2020) GenBank. *Nucleic Acids Research*, 48(D1), p. D84–D86. Available online: https://doi.org/10.1093/nar/gkz956.

Schiavinato, M., Bodrug-Schepers, A., Dohm, J.C. & Himmelbauer, H. (2021) Subgenome evolution in allotetraploid plants. *The Plant Journal: For Cell and Molecular Biology*, 106(3), p. 672–688. Available online: https://doi.org/10.1111/tpj.15190.

Schoenfelder, K.P. & Fox, D.T. (2015) The expanding implications of polyploidy. *The Journal of Cell Biology*, 209(4), p. 485–491. Available online: https://doi.org/10.1083/jcb.201502016.

Seah, S., Telleen, A.C. & Williamson, V.M. (2007) Introgressed and endogenous Mi-1 gene clusters in tomato differ by complex rearrangements in flanking sequences and show sequence exchange and diversifying selection among homologues. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, 114(7), p. 1289–1302. Available online: https://doi.org/10.1007/s00122-007-0519-z.

Sedlazeck, F.J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., von Haeseler, A. & Schatz, M.C. (2018) Accurate detection of complex structural variations using single-molecule sequencing. *Nature Methods*, 15(6), p. 461–468. Available online: https://doi.org/10.1038/s41592-018-0001-7.

Seehausen, O., Butlin, R.K., Keller, I., Wagner, C.E., Boughman, J.W., Hohenlohe, P.A., Peichel, C.L., Saetre, G.-P., Bank, C., Brännström, A., Brelsford, A., Clarkson, C.S., Eroukhmanoff, F., Feder, J.L., Fischer, M.C., Foote, A.D., Franchini, P., Jiggins, C.D., Jones, F.C., Lindholm, A.K., Lucek, K., Maan, M.E., Marques, D.A., Martin, S.H., Matthews, B., Meier, J.I., Möst, M., Nachman, M.W., Nonaka, E., Rennison, D.J., Schwarzer, J., Watson, E.T., Westram, A.M. & Widmer, A. (2014) Genomics and the origin of species. *Nature Reviews. Genetics*, 15(3), p. 176–192. Available online: https://doi.org/10.1038/nrg3644.

Sellers, G.S., Jeffares, D.C., Lawson, B., Prior, T. & Lunt, D.H. (2021) Identification of individual root-knot nematodes using low coverage long-read sequencing. *PloS One*, 16(12), p. e0253248. Available online: https://doi.org/10.1371/journal.pone.0253248.

Selmecki, A.M., Maruvka, Y.E., Richmond, P.A., Guillet, M., Shoresh, N., Sorenson, A.L., De, S., Kishony, R., Michor, F., Dowell, R. & Pellman, D. (2015) Polyploidy can drive rapid adaptation in yeast. *Nature*, 519(7543), p. 349–352. Available online: https://doi.org/10.1038/nature14187.

Session, A.M., Uno, Y., Kwon, T., Chapman, J.A., Toyoda, A., Takahashi, S., Fukui, A., Hikosaka, A., Suzuki, A., Kondo, M., van Heeringen, S.J., Quigley, I., Heinz, S., Ogino, H., Ochi, H., Hellsten, U., Lyons, J.B., Simakov, O., Putnam, N., Stites, J., Kuroki, Y., Tanaka, T., Michiue, T., Watanabe, M., Bogdanovic, O., Lister, R., Georgiou, G., Paranjpe, S.S., van Kruijsbergen, I., Shu, S., Carlson, J., Kinoshita, T., Ohta, Y., Mawaribuchi, S., Jenkins, J., Grimwood, J., Schmutz, J., Mitros, T., Mozaffari, S.V., Suzuki, Y., Haramoto, Y., Yamamoto, T.S., Takagi, C., Heald, R., Miller, K., Haudenschild, C., Kitzman, J., Nakayama, T., Izutsu, Y., Robert, J., Fortriede, J., Burns, K., Lotay, V., Karimi, K., Yasuoka, Y., Dichmann, D.S., Flajnik, M.F., Houston, D.W., Shendure, J., DuPasquier, L., Vize, P.D., Zorn, A.M., Ito, M., Marcotte, E.M., Wallingford, J.B., Ito, Y., Asashima, M., Ueno, N., Matsuda, Y., Veenstra, G.J.C., Fujiyama, A., Harland, R.M., Taira, M. & Rokhsar, D.S. (2016) Genome evolution in the allotetraploid frog Xenopus laevis. *Nature*, 538(7625), p. 336–343. Available online: https://doi.org/10.1038/nature19840.

Shafin, K., Pesout, T., Lorig-Roach, R., Haukness, M., Olsen, H.E., Bosworth, C., Armstrong, J., Tigyi, K., Maurer, N., Koren, S., Sedlazeck, F.J., Marschall, T., Mayes, S., Costa, V., Zook, J.M., Liu, K.J., Kilburn, D., Sorensen, M., Munson, K.M., Vollger, M.R., Monlong, J., Garrison, E., Eichler, E.E., Salama, S., Haussler, D., Green, R.E., Akeson, M., Phillippy, A., Miga, K.H., Carnevali, P., Jain, M. & Paten, B. (2020) Nanopore sequencing and the Shasta toolkit enable efficient de novo assembly of eleven human genomes. *Nature Biotechnology*, 38(9), p. 1044–1053. Available online: https://doi.org/10.1038/s41587-020-0503-6.

Shan, S., Gitzendanner, M.A., Boatwright, J.L., Spoelhof, J.P., Ethridge, C.L., Ji, L., Liu, X., Soltis, P.S., Schmitz, R.J. & Soltis, D.E. (2024) Genome-wide DNA methylation dynamics following recent polyploidy in the allotetraploid Tragopogon miscellus (Asteraceae). *The New Phytologist*, 242(3), p. 1363–1376. Available online: https://doi.org/10.1111/nph.19655.

Sharma, V., Hecker, N., Roscito, J.G., Foerster, L., Langer, B.E. & Hiller, M. (2018) A genomics approach reveals insights into the importance of gene losses for mammalian adaptations. *Nature Communications*, 9(1), p. 1215. Available online: https://doi.org/10.1038/s41467-018-03667-1.

Shen, Y., Li, W., Zeng, Y., Li, Z., Chen, Y., Zhang, J., Zhao, H., Feng, L., Ma, D., Mo, X., Ouyang, P., Huang, L., Wang, Z., Jiao, Y. & Wang, H.-B. (2022) Chromosome-level and haplotype-resolved genome provides insight into the tetraploid hybrid origin of patchouli. *Nature Communications*, 13(1), p. 1–15. Available online: https://doi.org/10.1038/s41467-022-31121-w.

Shilpa, Sharma, P., Thakur, V., Sharma, A., Rana, R.S. & Kumar, P. (2022) A status-quo review on management of root knot nematode in tomato. *The Journal of Horticultural Science & Biotechnology*, 97(4), p. 403–416. Available online: https://doi.org/10.1080/14620316.2022.2034531.

Shim, J., Kim, Y., Humphreys, G.I., Nardulli, A.M., Kosari, F., Vasmatzis, G., Taylor, W.R., Ahlquist, D.A., Myong, S. & Bashir, R. (2015) Nanopore-Based Assay for Detection of Methylation in Double-Stranded DNA Fragments. *ACS Nano*, 9(1), p. 290–300. Available online: https://doi.org/10.1021/nn5045596.

Shindo, T. (2010) *Investigating the role of papain-like cysteine protease RD21 in plant-pathogen interactions*. text.thesis.doctoral. Universität zu Köln. Available online: http://kups.ub.uni-koeln.de/id/eprint/3086 [Accessed 01/07/2024].

Shindo, T. & Van der Hoorn, R.A.L. (2008) Papain-like cysteine proteases: key players at molecular battlefields employed by both plants and their invaders. *Molecular Plant Pathology*, 9(1), p. 119–125. Available online: https://doi.org/10.1111/j.1364-3703.2007.00439.x.

Shu, D., Isozaki, Y., Zhang, X., Han, J. & Maruyama, S. (2014) Birth and early evolution of metazoans. *Gondwana Research*, 25(3), p. 884–895. Available online: https://doi.org/10.1016/j.gr.2013.09.001.

Shumate, A. & Salzberg, S.L. (2021) Liftoff: accurate mapping of gene annotations. *Bioinformatics* , 37(12), p. 1639–1643. Available online: https://doi.org/10.1093/bioinformatics/btaa1016.

Sikandar, A., Jia, L., Wu, H. & Yang, S. (2022) Meloidogyne enterolobii risk to agriculture, its present status and future prospective for management. *Frontiers in Plant Science*, 13, p. 1093657. Available online: https://doi.org/10.3389/fpls.2022.1093657.

Sikandar, A., Zhang, M.Y., Wang, Y.Y., Zhu, X.F., Liu, X.Y., Fan, H.Y., Xuan, Y.H., Chen, L.J. & Duan, Y.X. (2020) Meloidogyne incognita (root-knot nematode) a risk to agriculture. *Applied Ecology and Environmental Research*, 18(1). Available online: https://www.aloki.hu/pdf/1801_16791690.pdf.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. & Zdobnov, E.M. (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* , 31(19), p. 3210–3212. Available online: https://doi.org/10.1093/bioinformatics/btv351.

Singh, S., Singh, B. & Singh, A.P. (2015) Nematodes: A threat to sustainability of agriculture. *Procedia Environmental Sciences*, 29, p. 215–216. Available online: https://doi.org/10.1016/j.proenv.2015.07.270.

Sivashankari, S. & Shanmughavel, P. (2006) Functional annotation of hypothetical proteins - A review. *Bioinformation*, 1(8), p. 335–338. Available online: https://doi.org/10.6026/97320630001335.

Slater, G.S.C. & Birney, E. (2005) Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, 6(1), p. 31. Available online: https://doi.org/10.1186/1471-2105-6-31.

Smit, A.F.A., Hubley, R. & Green, P. (2015a) RepeatMasker Open-4.0. 2013--2015.

Smit, A.F.A., Hubley, R. & Green, P. (2015b) RepeatModeler Open-1.0. 2008--2015. *Seattle, USA: Institute for Systems Biology. Available from: Httpwww. Repeatmasker. Org, Last Accessed May*, 1, p. 2018.

Smith, M.D., Wertheim, J.O., Weaver, S., Murrell, B., Scheffler, K. & Kosakovsky Pond, S.L. (2015) Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Molecular Biology and Evolution*, 32(5), p. 1342–1353. Available online: https://doi.org/10.1093/molbev/msv022.

Soltani, T., Nejad, R.F., Ahmadi, A. & Fayazi, F. (2013) Chemical control of Root-Knot Nematode (Meloidogyne javanica) On Olive in the Greenhouse conditions. *Journal of Plant Pathology & Microbiology*, 4, p. 1–4. Available online: https://doi.org/10.4172/2157-7471.1000183.

Sommer, R.J. & Streit, A. (2011) Comparative genetics and genomics of nematodes: genome structure, development, and lifestyle. *Annual Review of Genetics*, 45, p. 1–20. Available online: https://doi.org/10.1146/annurev-genet-110410-132417.

Somvanshi, V.S., Dash, M., Bhat, C.G., Budhwar, R., Godwin, J., Shukla, R.N., Patrignani, A., Schlapbach, R. & Rao, U. (2021) An improved draft genome assembly of Meloidogyne graminicola IARI strain using long-read sequencing. *Gene*, 793(145748), p. 145748. Available online: https://doi.org/10.1016/j.gene.2021.145748.

Song, A., Su, J., Wang, H., Zhang, Z., Zhang, X., Van de Peer, Y., Chen, F., Fang, W., Guan, Z., Zhang, F., Wang, Z., Wang, L., Ding, B., Zhao, S., Ding, L., Liu, Y., Zhou, L., He, J., Jia, D., Zhang, J., Chen, C., Yu, Z., Sun, D., Jiang, J., Chen, S. & Chen, F. (2023) Analyses of a chromosome-scale genome assembly reveal the origin and evolution of cultivated chrysanthemum. *Nature Communications*, 14(1), p. 2021. Available online: https://doi.org/10.1038/s41467-023-37730-3.

Song, H., Lin, B., Huang, Q., Sun, T., Wang, W., Liao, J. & Zhuo, K. (2021) The Meloidogyne javanica effector Mj2G02 interferes with jasmonic acid signalling to suppress cell death and promote parasitism in Arabidopsis. *Molecular Plant Pathology*, 22(10), p. 1288–1301. Available online: https://doi.org/10.1111/mpp.13111.

Song, Q. & Chen, Z.J. (2015) Epigenetic and developmental regulation in plant polyploids. *Current Opinion in Plant Biology*, 24, p. 101–109. Available online: https://doi.org/10.1016/j.pbi.2015.02.007.

Spjuth, O., Bongcam-Rudloff, E., Hernández, G.C., Forer, L., Giovacchini, M., Guimera, R.V., Kallio, A., Korpelainen, E., Kańduła, M.M., Krachunov, M., Kreil, D.P., Kulev, O., Łabaj, P.P., Lampa, S., Pireddu, L., Schönherr, S., Siretskiy, A. & Vassilev, D. (2015) Experiences with workflows for automating data-intensive bioinformatics. *Biology Direct*, 10(1), p. 43. Available online: https://doi.org/10.1186/s13062-015-0071-8.

Stebbins, G. (1947) Types of polyploids; their classification and significance. *Advances in Genetics*, 1, p. 403–429. Available online: https://doi.org/10.1016/s0065-2660(08)60490-3.

Stephenson-Gussinye, A. & Furlan-Magaril, M. (2023) Chromosome conformation capture technologies as tools to detect structural variations and their repercussion in chromatin 3D configuration. *Frontiers in Cell and Developmental Biology*, 11. Available online: https://doi.org/10.3389/fcell.2023.1219968.

Stevens, L., Martínez-Ugalde, I., King, E., Wagah, M., Absolon, D., Bancroft, R., Gonzalez de la Rosa, P., Hall, J.L., Kieninger, M., Kloch, A., Pelan, S., Robertson, E., Pedersen, A.B., Abreu-Goodger, C., Buck, A.H. & Blaxter, M. (2023) Ancient diversity in host-parasite interaction genes in a model parasitic nematode. *Nature Communications*, 14(1), p. 7776. Available online: https://doi.org/10.1038/s41467-023-43556-w.

Stodden, V., McNutt, M., Bailey, D., Deelman, E., Gil, Y., Hanson, B., Heroux, M., Ioannidis, J. & Taufer, M. (2016) Enhancing reproducibility for computational methods. *Science*, 354, p. 1240–1241. Available online: https://doi.org/10.1126/science.aah6168.

Stothard, P. (2000) The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *BioTechniques*, (28), p. 1102–1104.

Stothard, P. & Pilgrim, D. (2003) Sex-determination gene and pathway evolution in nematodes. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 25(3), p. 221–231. Available online: https://doi.org/10.1002/bies.10239.

Subbotin, S.A., Palomares Rius, J.E. & Castillo, P. (2021) *Systematics of Root-Knot Nematodes (Nematoda: Meloidogynidae)*. Leiden, Netherlands: Brill (Nematology Monographs and Perspectives, 14). Available online: https://books.google.com/books?hl=en&lr=&id=xs87EAAAQBAJ&oi=fnd&pg=PR3&ots=g2aY8hk6QX&sig=duuEag_HdDVR4EvDB5Jjsgr100U.

Subbotin, S.A., Rius, J.E.P. & Castillo, P. (2021) *Systematics of Root-knot Nematodes (Nematoda: Meloidogynidae)*. BRILL. Available online: https://play.google.com/store/books/details?id=xs87EAAAQBAJ.

Suomalainen, E. (1950) Parthenogenesis in animals. *Advances in Genetics*, 3, p. 193–253. Available online: https://www.sciencedirect.com/science/article/pii/S0065266008600863.

Susič, N., Koutsovoulos, G.D., Riccio, C., Danchin, E.G.J., Blaxter, M.L., Lunt, D.H., Strajnar, P., Širca, S., Urek, G. & Stare, B.G. (2020) Genome sequence of the root-knot nematode Meloidogyne luci. *Journal of Nematology*, 52, p. 1–5. Available online: https://doi.org/10.21307/jofnem-2020-025.

Szitenberg, A., Salazar-Jaramillo, L., Blok, V.C., Laetsch, D.R., Joseph, S., Williamson, V.M., Blaxter, M.L. & Lunt, D.H. (2017) Comparative Genomics of Apomictic Root-Knot Nematodes: Hybridization, Ploidy, and Dynamic Genome Change. *Genome Biology and Evolution*, 9(10), p. 2844–2861. Available online: https://doi.org/10.1093/gbe/evx201.

Tachack, E.B., Oviedo-Socarrás, T., Pastrana, M.O., Pérez-Cogollo, L.C., Benavides, Y.H., Pinto, C.R. & Garay, O.V. (2022) Status of gastrointestinal nematode infections and associated epidemiological factors in sheep from Córdoba, Colombia. *Tropical Animal Health and Production*, 54(3), p. 171. Available online: https://doi.org/10.1007/s11250-022-03170-2.

Talavera-Rubia, M., Vela-Delgado, M.D. & Verdejo-Lucas, S. (2022) A cost-benefit analysis of soil disinfestation methods against root-knot nematodes in Mediterranean intensive horticulture. *Plants*, 11(20), p. 2774. Available online: https://doi.org/10.3390/plants11202774.

Tang, H., Bowers, J.E., Wang, X., Ming, R., Alam, M. & Paterson, A.H. (2008) Synteny and collinearity in plant genomes. *Science*, 320(5875), p. 486–488. Available online: https://doi.org/10.1126/science.1153917.

Tang, H., Krishnakumar, V. & Li, J. (2015) *Jcvi: JCVI utility libraries*. Available online: https://doi.org/10.5281/zenodo.31631.

Tang, R., Yu, Z. & Li, J. (2023) KINN: An alignment-free accurate phylogeny reconstruction method based on inner distance distributions of k-mer pairs in biological sequences. *Molecular Phylogenetics and Evolution*, 179(107662), p. 107662. Available online: https://doi.org/10.1016/j.ympev.2022.107662.

Tarailo-Graovac, M. & Chen, N. (2009) Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al.]*, Chapter 4, p. Unit 4.10. Available online: https://doi.org/10.1002/0471250953.bi0410s25.

Tariq-Khan, M., Mukhtar, T., Munir, A., Hallmann, J. & Heuer, H. (2020) Comprehensive report on the prevalence of root-knot nematodes in the Poonch division of Azad Jammu and Kashmir, Pakistan. *Phytopathologische Zeitschrift. Journal of Phytopathology*, 168(6), p. 322–336. Available online: https://doi.org/10.1111/jph.12895.

Thiermann, F., Vismann, B. & Giere, O. (2000) Sulphide tolerance of the marine nematode Oncholaimus campylocercoides-a result of internal sulphur formation?. *Marine Ecology Progress Series*, 193, p. 251–259. Available online: https://doi.org/10.3354/meps193251.

Thorvaldsdóttir, H., Robinson, J.T. & Mesirov, J.P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, 14(2), p. 178–192. Available online: https://doi.org/10.1093/bib/bbs017.

Thrash, A., Hoffmann, F. & Perkins, A. (2020) Toward a more holistic method of genome assembly assessment. *BMC Bioinformatics*, 21(Suppl 4), p. 249. Available online: https://doi.org/10.1186/s12859-020-3382-4.

Tian, G., Li, G., Liu, Y., Liu, Q., Wang, Y., Xia, G. & Wang, M. (2020) Polyploidization is accompanied by synonymous codon usage bias in the chloroplast genomes of both cotton and wheat. *PloS One*, 15(11), p. e0242624. Available online: https://doi.org/10.1371/journal.pone.0242624.

Tian, G., Xiao, G., Wu, T., Zhou, J., Xu, W., Wang, Y., Xia, G. & Wang, M. (2022) Alteration of synonymous codon usage bias accompanies polyploidization in wheat. *Frontiers in Genetics*, 13, p. 979902. Available online: https://doi.org/10.3389/fgene.2022.979902.

Tichkule, S., Jex, A.R., van Oosterhout, C., Sannella, A.R., Krumkamp, R., Aldrich, C., Maiga-Ascofare, O., Dekker, D., Lamshöft, M., Mbwana, J., Rakotozandrindrainy, N., Borrmann, S., Thye, T., Schuldt, K., Winter, D., Kremsner, P.G., Oppong, K., Manouana, P., Mbong, M., Gesase, S., Minja, D.T.R., Mueller, I., Bahlo, M., Nader, J., May, J., Rakotozandrindrain, R., Adegnika, A.A., Lusingu, J.P.A., Amuasi, J., Eibach, D. & Caccio, S.M. (2021) Comparative genomics revealed adaptive admixture in Cryptosporidium hominis in Africa. *Microbial Genomics*, 7(1). Available online: https://doi.org/10.1099/mgen.0.000493.

Tigano, A., Colella, J.P. & MacManes, M.D. (2020) Comparative and population genomics approaches reveal the basis of adaptation to deserts in a small rodent. *Molecular Ecology*, 29(7), p. 1300–1314. Available online: https://doi.org/10.1111/mec.15401.

de Tomás, C. & Vicient, C.M. (2023) The genomic shock hypothesis: Genetic and epigenetic alterations of transposable elements after interspecific hybridization in plants. *Epigenomes*, 8(1). Available online: https://doi.org/10.3390/epigenomes8010002.

Toups, M.A., Vicoso, B. & Pannell, J.R. (2022) Dioecy and chromosomal sex determination are maintained through allopolyploid speciation in the plant genus Mercurialis. *PLoS Genetics*, 18(7), p. e1010226. Available online: https://doi.org/10.1371/journal.pgen.1010226.

Trappe, K., Marschall, T. & Renard, B.Y. (2016) Detecting horizontal gene transfer by mapping sequencing reads across species boundaries. *Bioinformatics*, 32(17), p. i595–i604. Available online: https://doi.org/10.1093/bioinformatics/btw423.

Triantaphyllou, A.C. (1963) POLYPLOIDY AND PARTHENOGENESIS IN THE ROOT-KNOT NEMATODE MELOIDOGYNE ARENARIA. *Journal of Morphology*, 113, p. 489–499. Available online: https://doi.org/10.1002/jmor.1051130309.

Triantaphyllou, A.C. (1981) Oogenesis and the Chromosomes of the Parthenogenic Root-knot Nematode Meloidogyne incognita. *Journal of Nematology*, 13(2), p. 95–104. Available online: https://www.ncbi.nlm.nih.gov/pubmed/19300730.

Triantaphyllou, A.C. (1985) Gametogenesis and the Chromosomes of Meloidogyne nataliei: Not Typical of Other Root-knot Nematodes. *Journal of Nematology*, 17(1), p. 1–5. Available online: https://www.ncbi.nlm.nih.gov/pubmed/19294050.

Trudgill, D.L. & Blok, V.C. (2001) Apomictic, polyphagous root-knot nematodes: exceptionally successful and damaging biotrophic root pathogens. *Annual Review of Phytopathology*, 39, p. 53–77. Available online: https://doi.org/10.1146/annurev.phyto.39.1.53.

Tumescheit, C., Firth, A.E. & Brown, K. (2022) CIAlign: A highly customisable command line tool to clean, interpret and visualise multiple sequence alignments. *PeerJ*, 10, p. e12983. Available online: https://doi.org/10.7717/peerj.12983.

Turkyilmaz-van der Velden, Y., Dintzner, N. & Teperek, M. (2020) Reproducibility starts from you today. *Patterns (New York, N.Y.)*, 1(6), p. 100099. Available online: https://doi.org/10.1016/j.patter.2020.100099.

Tzortzakakis, E.A., Adam, M.A.M., Blok, V.C., Paraskevopoulos, C. & Bourtzis, K. (2005) Occurrence of Resistance-breaking Populations of Root-knot Nematodes on Tomato in Greece. *European Journal of Plant Pathology / European Foundation for Plant Pathology*, 113(1), p. 101–105. Available online: https://doi.org/10.1007/s10658-005-1228-6.

Urek, G. (2013) Biological Control of Root-Knot Nematodes (Meloidogyne spp.): Microbes against the Pests. *Acta Agriculturae Slovenica* [Preprint]. Available online: https://doi.org/10.2478/ACAS-2013-0022.

VanBuren, R., Man Wai, C., Wang, X., Pardo, J., Yocca, A.E., Wang, H., Chaluvadi, S.R., Han, G., Bryant, D., Edger, P.P., Messing, J., Sorrells, M.E., Mockler, T.C., Bennetzen, J.L. & Michael, T.P. (2020) Exceptional subgenome stability and functional divergence in the allotetraploid Ethiopian cereal teff. *Nature Communications*, 11(1). Available online: https://doi.org/10.1038/s41467-020-14724-z.

Van Etten, J., Stephens, T.G. & Bhattacharya, D. (2023) A k-mer-based approach for phylogenetic classification of taxa in environmental genomic data. *Systematic Biology*, 72(5), p. 1101–1118. Available online: https://doi.org/10.1093/sysbio/syad037.

Van Goor, J., Kanzaki, N. & Woodruff, G. (2023) How to be a fig nematode. *Acta Oecologica (Montrouge, France)*, 119(103916), p. 103916. Available online: https://doi.org/10.1016/j.actao.2023.103916.

Vicient, C.M. & Casacuberta, J.M. (2017) Impact of transposable elements on polyploid plant genomes. *Annals of Botany*, 120(2), p. 195–207. Available online: https://doi.org/10.1093/aob/mcx078.

Vrain, T.C., Barker, K.R. & Holtzman, G.I. (1978) Influence of low temperature on rate of development of Meloidogyne incognita and M. hapla larvae. *Journal of Nematology*, 10(2), p. 166–171. Available online: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2617882/pdf/166.pdf.

Walker, M.D. & Zunt, J.R. (2005) Neuroparasitic infections: nematodes. *Seminars in Neurology*, 25(3), p. 252–261. Available online: https://doi.org/10.1055/s-2005-917662.

Walve, R. (2023) Improving Contiguity and Accuracy in Genome Assembly. Available online: https://helda.helsinki.fi/server/api/core/bitstreams/61d415af-2124-4011-a256-3683aa59a18b/content.

Wang, P., Meng, F., Moore, B.M. & Shiu, S.-H. (2021) Impact of short-read sequencing on the misassembly of a plant genome. *BMC Genomics*, 22(1), p. 99. Available online: https://doi.org/10.1186/s12864-021-07397-5.

Wang, P. & Wang, F. (2022) A proposed metric set for evaluation of genome assembly quality. *Trends in Genetics: TIG* [Preprint]. Available online: https://doi.org/10.1016/j.tig.2022.10.005.

Wang, X.-R., Moreno, Y.A., Wu, H.-R., Ma, C., Li, Y.-F., Zhang, J.-A., Yang, C., Sun, S., Ma, W.-J. & Geary, T.G. (2012) Proteomic profiles of soluble proteins from the esophageal gland in female Meloidogyne incognita. *International Journal for Parasitology*, 42(13-14), p. 1177–1183. Available online: https://doi.org/10.1016/j.ijpara.2012.10.008.

Wang, X., You, X., Langer, J.D., Hou, J., Rupprecht, F., Vlatkovic, I., Quedenau, C., Tushev, G., Epstein, I., Schaefke, B., Sun, W., Fang, L., Li, G., Hu, Y., Schuman, E.M. & Chen, W. (2019) Full-length transcriptome reconstruction reveals a large diversity of RNA and protein isoforms in rat hippocampus. *Nature Communications*, 10(1), p. 5009. Available online: https://doi.org/10.1038/s41467-019-13037-0.

Wang, Y., Tang, H., Debarry, J.D., Tan, X., Li, J., Wang, X., Lee, T.-H., Jin, H., Marler, B., Guo, H., Kissinger, J.C. & Paterson, A.H. (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research*, 40(7), p. e49. Available online: https://doi.org/10.1093/nar/gkr1293.

Wang, Y., Yu, J., Jiang, M., Lei, W., Zhang, X. & Tang, H. (2023) Sequencing and Assembly of Polyploid Genomes: Methods and Protocols. In Van de Peer, Y. (ed.) *Polyploidy*. New York, NY: Springer US (Methods in Molecular Biology), p. 429–458. Available online: https://doi.org/10.1007/978-1-0716-2561-3_23.

Wang, Z., Yang, J., Cheng, F., Li, P., Xin, X., Wang, W., Yu, Y., Zhang, D., Zhao, X., Yu, S., Zhang, F., Dong, Y. & Su, T. (2022) Subgenome dominance and its evolutionary implications in crop domestication and breeding. *Horticulture Research*, 9, p. uhac090. Available online: https://doi.org/10.1093/hr/uhac090.

Weissenbach, J. (2016) The rise of genomics. *Comptes Rendus Biologies*, 339(7-8), p. 231–239. Available online: https://doi.org/10.1016/j.crvi.2016.05.002.

Wendel, J.F. (2000) Genome evolution in polyploids. *Plant Molecular Biology*, 42(1), p. 225–249. Available online: https://www.ncbi.nlm.nih.gov/pubmed/10688139.

Wenger, A.M., Peluso, P., Rowell, W.J., Chang, P.-C., Hall, R.J., Concepcion, G.T., Ebler, J., Fungtammasan, A., Kolesnikov, A., Olson, N.D., Töpfer, A., Alonge, M., Mahmoud, M., Qian, Y., Chin, C.-S., Phillippy, A.M., Schatz, M.C., Myers, G., DePristo, M.A., Ruan, J., Marschall, T., Sedlazeck, F.J., Zook, J.M., Li, H., Koren, S., Carroll, A., Rank, D.R. & Hunkapiller, M.W. (2019) Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nature Biotechnology*, 37(10), p. 1155–1162. Available online: https://doi.org/10.1038/s41587-019-0217-9.

Wesemael, W., Viaene, N. & Moens, M. (2011) Root-knot nematodes (Meloidogyne spp.) in Europe. *Nematology: International Journal of Fundamental and Applied Nematological Research*, 13(1), p. 3–16. Available online: https://doi.org/10.1163/138855410X526831.

Westerdahl, B.B. (2018) Evaluation of trap cropping for management of root-knot nematode on annual crops. In *IX International Symposium on Soil and Substrate Disinfestation 1270*. actahort.org, p. 141–146. Available online: https://www.actahort.org/books/1270/1270_15.htm.

Wharton, D.A. (1986a) Life Cycle. In *A Functional Biology of Nematodes*. Boston, MA: Springer US, p. 118–148. Available online: https://doi.org/10.1007/978-1-4615-8516-9_6.

Wharton, D.A. (1986b) Reproductive Biology. In *A Functional Biology of Nematodes*. Boston, MA: Springer US, p. 60–87. Available online: https://doi.org/10.1007/978-1-4615-8516-9_4.

Whibley, A., Kelley, J.L. & Narum, S.R. (2021) The changing face of genome assemblies: Guidance on achieving high-quality reference genomes. *Molecular Ecology Resources*, 21(3), p. 641–652. Available online: https://doi.org/10.1111/1755-0998.13312.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J. & Mons, B. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, p. 160018. Available online: https://doi.org/10.1038/sdata.2016.18.

Williamson, V.M. (1998) Root-knot nematode resistance genes in tomato and their potential for future use. *Annual Review of Phytopathology*, 36, p. 277–293. Available online: https://doi.org/10.1146/annurev.phyto.36.1.277.

Williamson, V.M. & Gleason, C.A. (2003) Plant–nematode interactions. *Current Opinion in Plant Biology*, 6(4), p. 327–333. Available online: https://doi.org/10.1016/S1369-5266(03)00059-1.

Williamson, V.M. & Hussey, R.S. (1996) Nematode pathogenesis and resistance in plants. *The Plant Cell*, 8(10), p. 1735–1745. Available online: https://doi.org/10.1105/tpc.8.10.1735.

Williamson, V.M. & Kumar, A. (2006) Nematode resistance in plants: the battle underground. *Trends in Genetics: TIG*, 22(7), p. 396–403. Available online: https://doi.org/10.1016/j.tig.2006.05.003.

Winter, M. (2022) *Asmapp: ASMAPP assembly appraisal workflow. Built in snakemake, ASMAPP performs many basic and intermediate assembly appraisal tasks*. Github. Available online: https://github.com/mrmrwinter/asmapp [Accessed 18/10/2022].

Winter, M.R. (2020) *Comparative genomics of Meloidogyne haplanaria*. Edited by D.H. Lunt. Masters by Research. University of Hull. Available online: https://hull-repository.worktribe.com/output/4223716.

Winter, M.R., Taranto, A.P., Yimer, H.Z., Blundell, A.C., Siddique, S., Williamson, V.M. & Lunt, D.H. (2023) Phased chromosome-scale genome assembly of an asexual, allopolyploid root-knot nematode reveals complex subgenomic structure. *bioRxiv*. Available online: https://doi.org/10.1101/2023.03.20.533402.

Winter, M.R., Taranto, A.P., Yimer, H.Z., Coomer Blundell, A., Siddique, S., Williamson, V.M. & Lunt, D.H. (2024) Phased chromosome-scale genome assembly of an asexual, allopolyploid root-knot nematode reveals complex subgenomic structure. *PloS One*, 19(6), p. e0302506. Available online: https://doi.org/10.1371/journal.pone.0302506.

Wolfe, K.H. (2001) Yesterday's polyploids and the mystery of diploidization. *Nature Reviews. Genetics*, 2(5), p. 333–341. Available online: https://doi.org/10.1038/35072009.

Wratten, L., Wilm, A. & Göke, J. (2021) Reproducible, scalable, and shareable analysis pipelines with bioinformatics workflow managers. *Nature Methods*, 18(10), p. 1161–1168. Available online: https://doi.org/10.1038/s41592-021-01254-9.

Wright, C.J., Stevens, L., Mackintosh, A., Lawniczak, M. & Blaxter, M. (2024) Comparative genomics reveals the dynamics of chromosome evolution in Lepidoptera. *Nature Ecology & Evolution*, 8(4), p. 777–790. Available online: https://doi.org/10.1038/s41559-024-02329-4.

Wyss, U. & Grundler, F.M.W. (1992) Feeding behavior of sedentary plant parasitic nematodes. *Netherlands Journal of Plant Pathology* [Preprint]. Available online: https://doi.org/10.1007/BF01974483.

Xiang, N., Lawrence, K.S. & Donald, P.A. (2018) Biological control potential of plant growth-promoting rhizobacteria suppression ofMeloidogyne incognitaon cotton andHeterodera glycineson soybean: A review. *Journal of Phytopathology (1986)*, 166(7-8), p. 449–458. Available online: https://doi.org/10.1111/jph.12712.

Yaghoobi, J., Kaloshian, I., Wen, Y. & Williamson, V.M. (1995) Mapping a new nematode resistance locus in Lycopersicon peruvianum. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, 91(3), p. 457–464. Available online: https://doi.org/10.1007/BF00222973.

Yamauchi, A., Hosokawa, A., Nagata, H. & Shimoda, M. (2004) Triploid bridge and role of parthenogenesis in the evolution of autopolyploidy. *The American Naturalist*, 164(1), p. 101–112. Available online: https://doi.org/10.1086/421356.

Yandell, M. & Ence, D. (2012) A beginner's guide to eukaryotic genome annotation. *Nature Reviews. Genetics*, 13(5), p. 329–342. Available online: https://doi.org/10.1038/nrg3174.

Yang, J., Liu, D., Wang, X., Ji, C., Cheng, F., Liu, B., Hu, Z., Chen, S., Pental, D., Ju, Y., Yao, P., Li, X., Xie, K., Zhang, J., Wang, J., Liu, F., Ma, W., Shopan, J., Zheng, H., Mackenzie, S.A. & Zhang, M. (2016) The genome sequence of allopolyploid Brassica juncea and analysis of differential homoeolog gene expression influencing selection. *Nature Genetics*, 48(10), p. 1225–1232. Available online: https://doi.org/10.1038/ng.3657.

Yuen, Z.W.-S., Srivastava, A., Daniel, R., McNevin, D., Jack, C. & Eyras, E. (2021) Systematic benchmarking of tools for CpG methylation detection from nanopore sequencing. *Nature Communications*, 12(1), p. 3438. Available online: https://doi.org/10.1038/s41467-021-23778-6.

Yu, G., Smith, D.K., Zhu, H., Guan, Y. & Lam, T.T. (2017) Ggtree : an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution. Edited by G. McInerny, 8(1), p. 28–36. Available online: https://doi.org/10.1111/2041-210X.12628.

Yu, S.-T., Zhao, R., Sun, X.-Q., Hou, M.-X., Cao, Y.-M., Zhang, J., Chen, Y.-J., Wang, K.-K., Zhang, Y., Li, J.-T. & Wang, Q. (2024) DNA Methylation and Chromatin Accessibility Impact Subgenome Expression Dominance in the Common Carp (Cyprinus carpio). *International Journal of Molecular Sciences*, 25(3). Available online: https://doi.org/10.3390/ijms25031635.

Zdobnov, E.M., Kuznetsov, D., Tegenfeldt, F., Manni, M., Berkeley, M. & Kriventseva, E.V. (2021) OrthoDB in 2020: evolutionary and functional annotations of orthologs. *Nucleic Acids Research*, 49(D1), p. D389–D393. Available online: https://doi.org/10.1093/nar/gkaa1009.

Zhang, K., Wang, X. & Cheng, F. (2019) Plant Polyploidy: Origin, Evolution, and Its Influence on Crop Domestication. *Horticultural Plant Journal*, 5(6), p. 231–239. Available online: https://doi.org/10.1016/j.hpj.2019.11.003.

Zhang, X., Wu, R., Wang, Y., Yu, J. & Tang, H. (2020) Unzipping haplotypes in diploid and polyploid genomes. *Computational and Structural Biotechnology Journal*, 18, p. 66–72. Available online: https://doi.org/10.1016/j.csbj.2019.11.011.

Zheng, Y., Yang, D., Rong, J., Chen, L., Zhu, Q., He, T., Chen, L., Ye, J., Fan, L., Gao, Y., Zhang, H. & Gu, L. (2022) Allele-aware chromosome-scale assembly of the allopolyploid genome of hexaploid Ma bamboo (Dendrocalamus latiflorus Munro). *Journal of Integrative Plant Biology*, 64(3), p. 649–670. Available online: https://doi.org/10.1111/jipb.13217.

Zhou, L., Feng, T., Xu, S., Gao, F., Lam, T.T., Wang, Q., Wu, T., Huang, H., Zhan, L., Li, L., Guan, Y., Dai, Z. & Yu, G. (2022) Ggmsa: a visual exploration tool for multiple sequence alignment and associated data. *Briefings in Bioinformatics*, 23(4). Available online: https://doi.org/10.1093/bib/bbac222.

Zhou, L. & Gui, J. (2017) Natural and artificial polyploids in aquaculture. *Aquaculture and Fisheries*, 2(3), p. 103–111. Available online: https://doi.org/10.1016/j.aaf.2017.04.003.

Ziemann, M., Poulain, P. & Bora, A. (2023) The five pillars of computational reproducibility: bioinformatics and beyond. *Briefings in Bioinformatics*, 24. Available online: https://doi.org/10.1093/bib/bbad375.

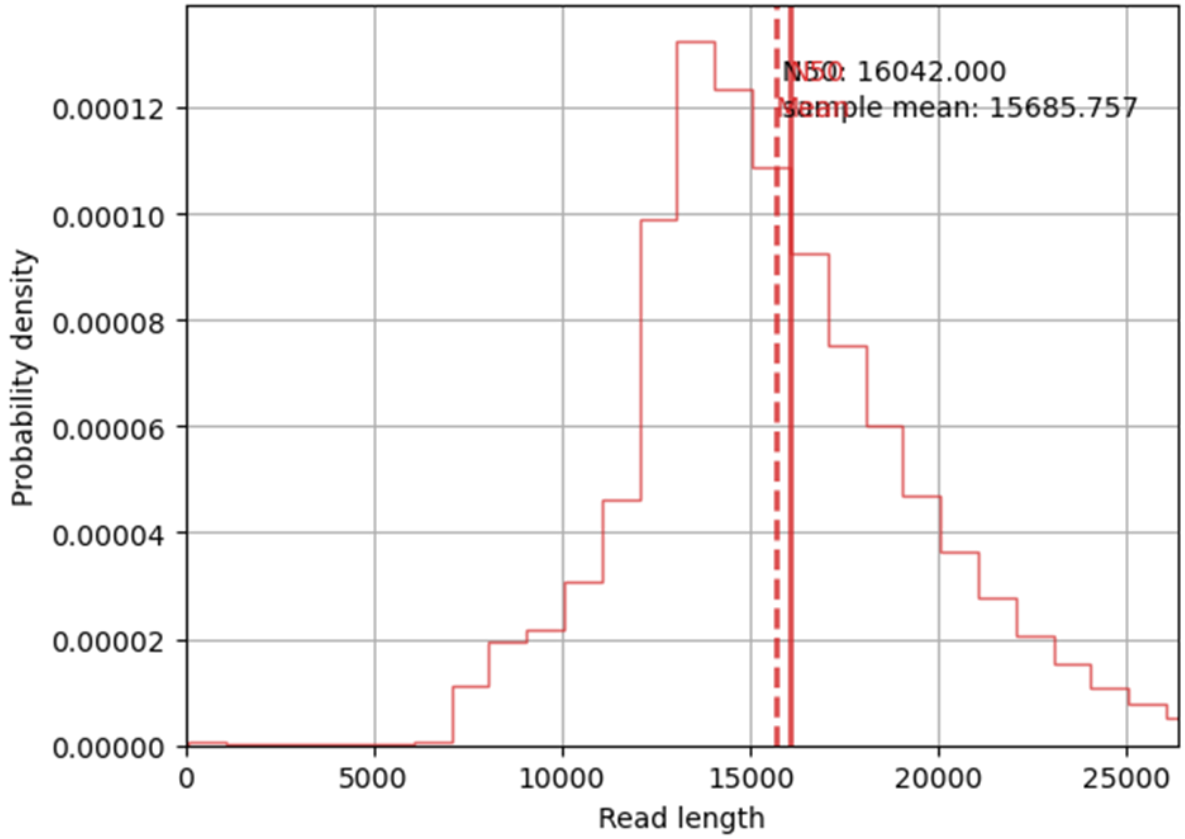
Zimin, A.V. & Salzberg, S.L. (2022) The SAMBA tool uses long reads to improve the contiguity of genome assemblies. *PLoS Computational Biology*, 18(2), p. e1009860. Available online: https://doi.org/10.1371/journal.pcbi.1009860.

Zörgö, E., Chwialkowska, K., Gjuvsland, A.B., Garré, E., Sunnerhagen, P., Liti, G., Blomberg, A., Omholt, S.W. & Warringer, J. (2013) Ancient evolutionary trade-offs between yeast ploidy states. *PLoS Genetics*, 9(3), p. e1003388. Available online: https://doi.org/10.1371/journal.pgen.1003388.

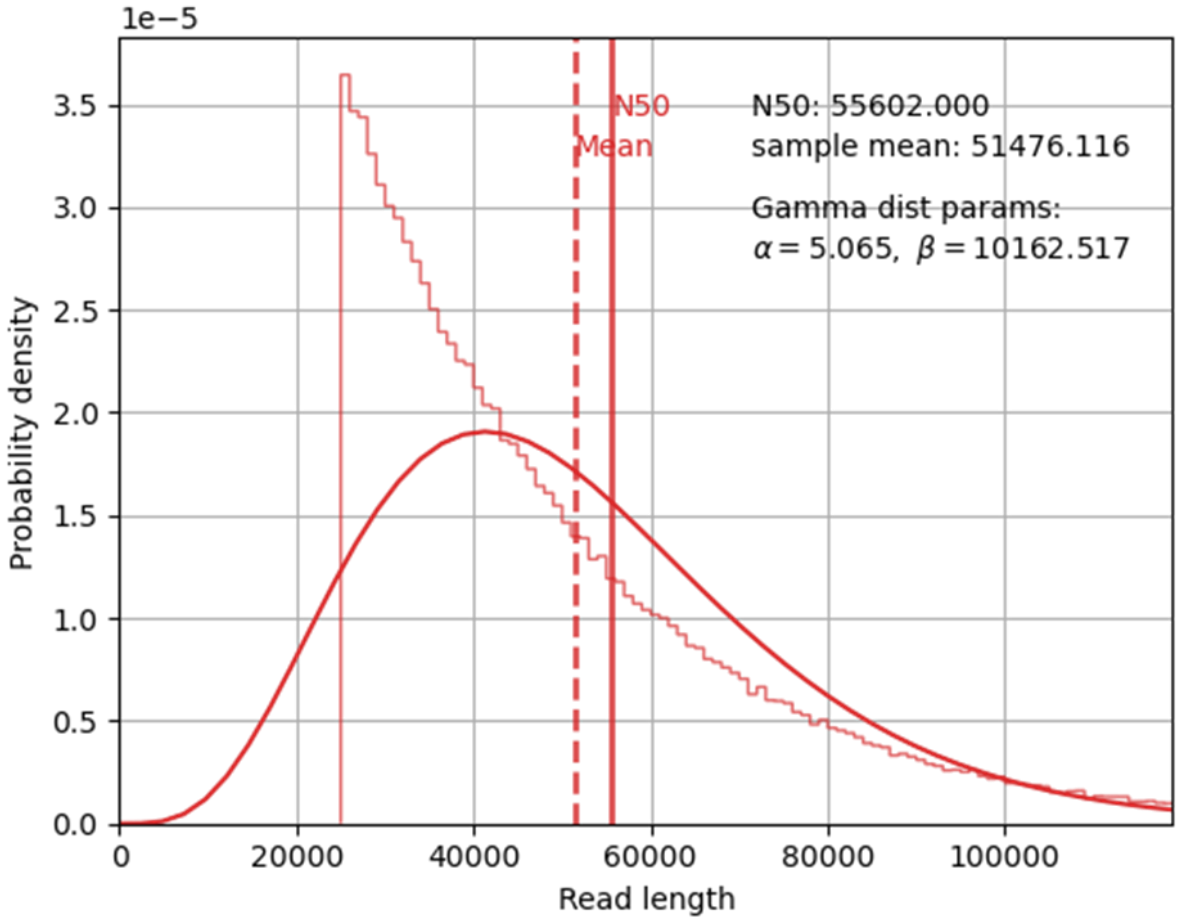
Zulkower, V. & Rosser, S. (2020) DNA Features Viewer: a sequence annotation formatting and plotting library for Python. *Bioinformatics* , 36(15), p. 4350–4352. Available online: https://doi.org/10.1093/bioinformatics/btaa213.

7.0 Supplementary

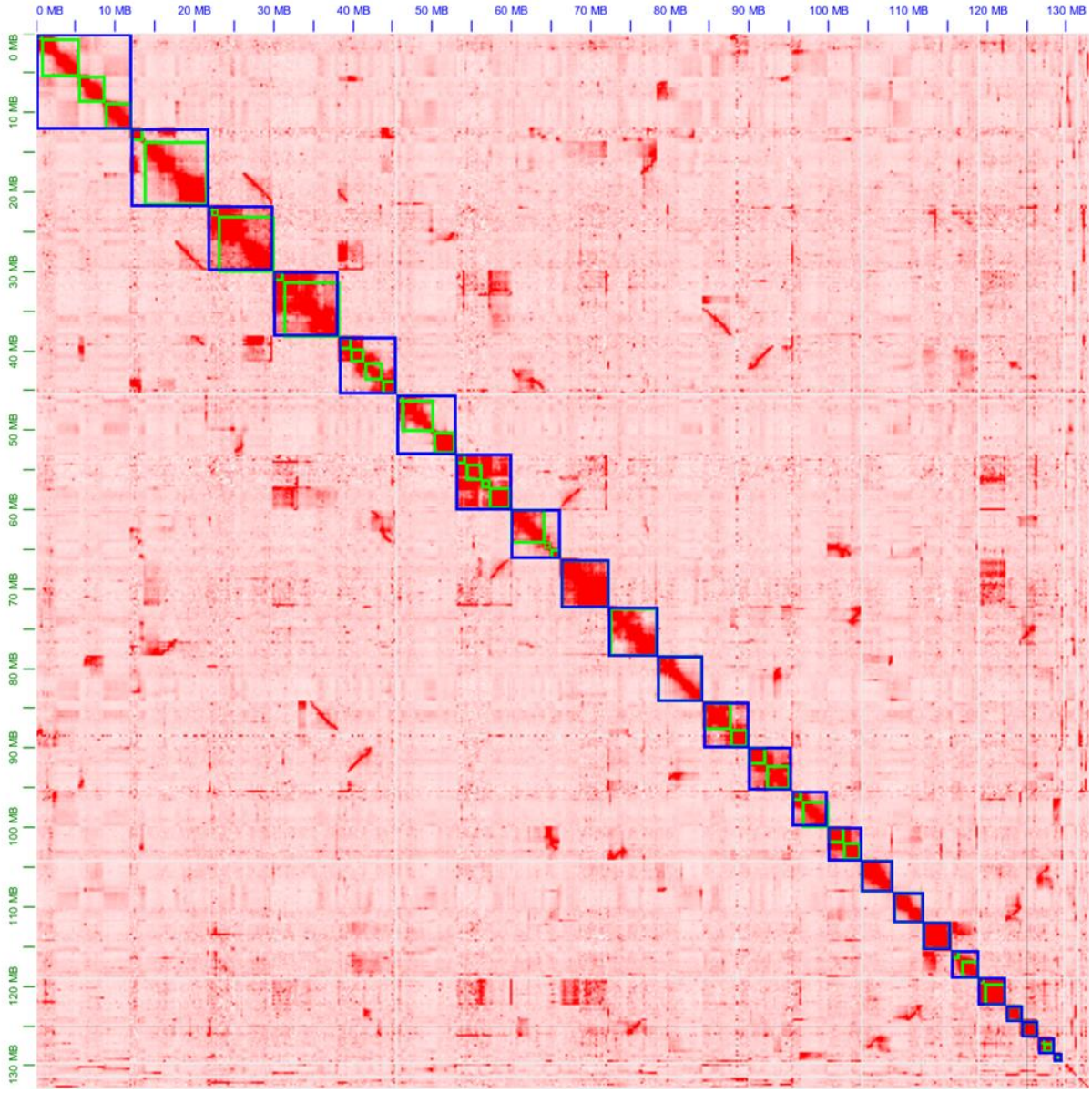
7.1 Supplementary Figures



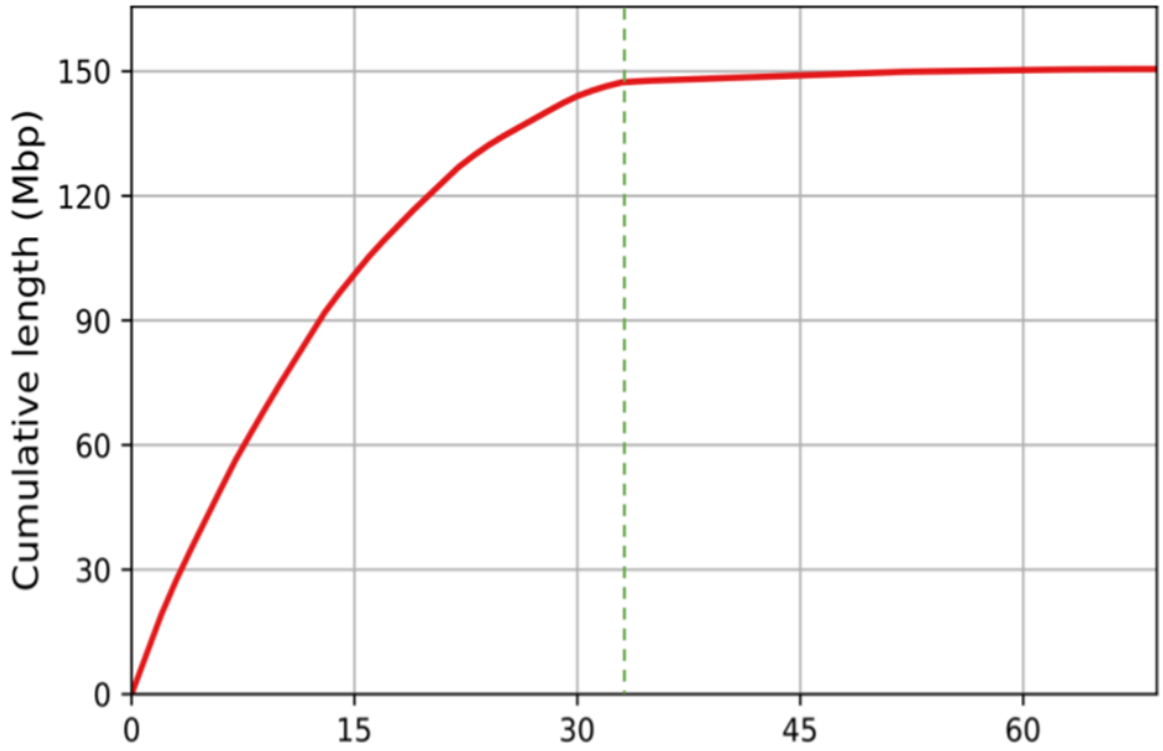
Supplementary Figure 2.1: Length and quality statistics of concatenated and quality controlled PacBio HiFi libraries. X axis is read length. Y axis is probability density.



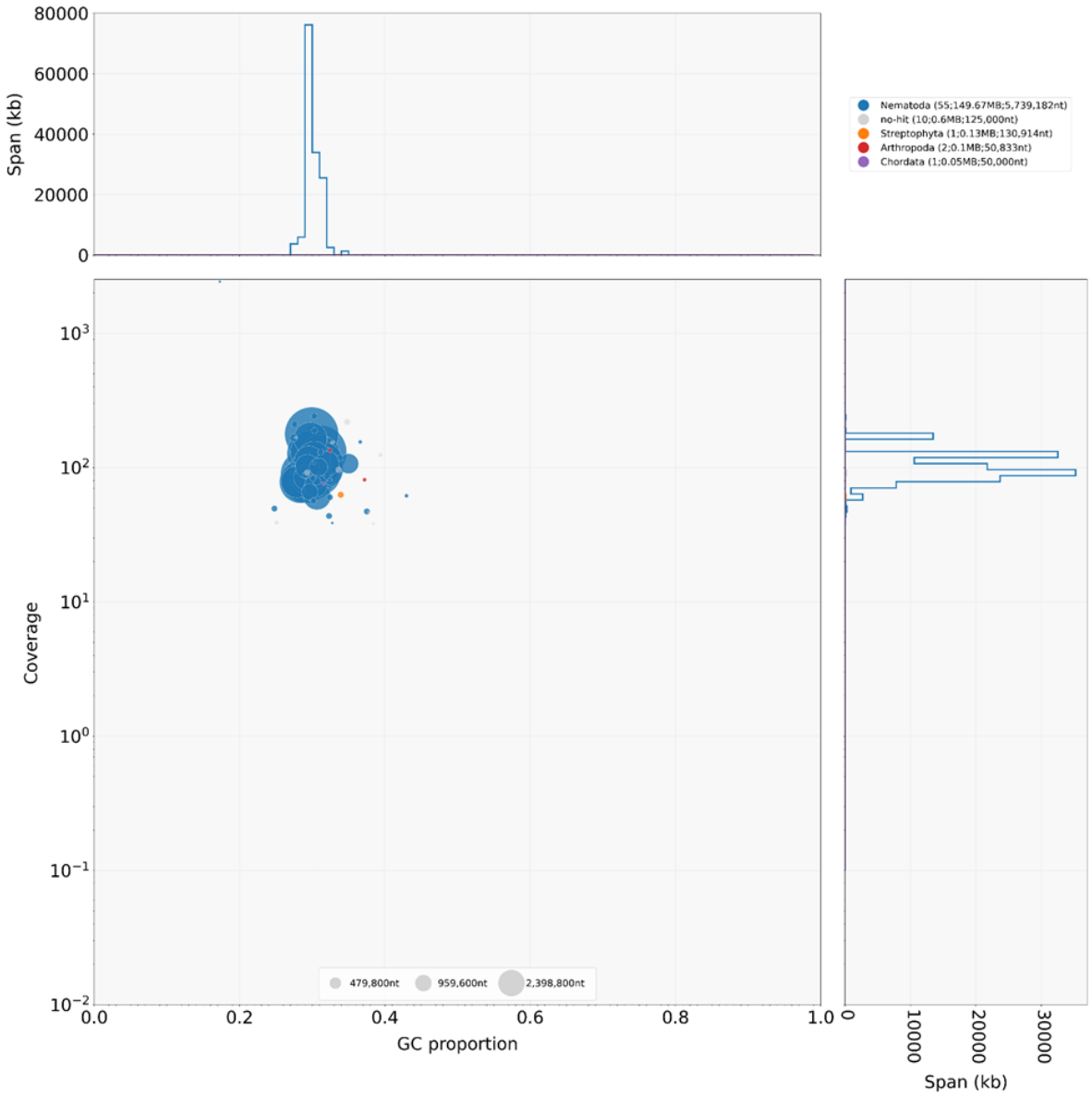
Supplementary Figure 2.2: Length and quality statistics of concatenated and quality controlled Oxford Nanopore libraries. X axis is read length. Y axis is probability density.



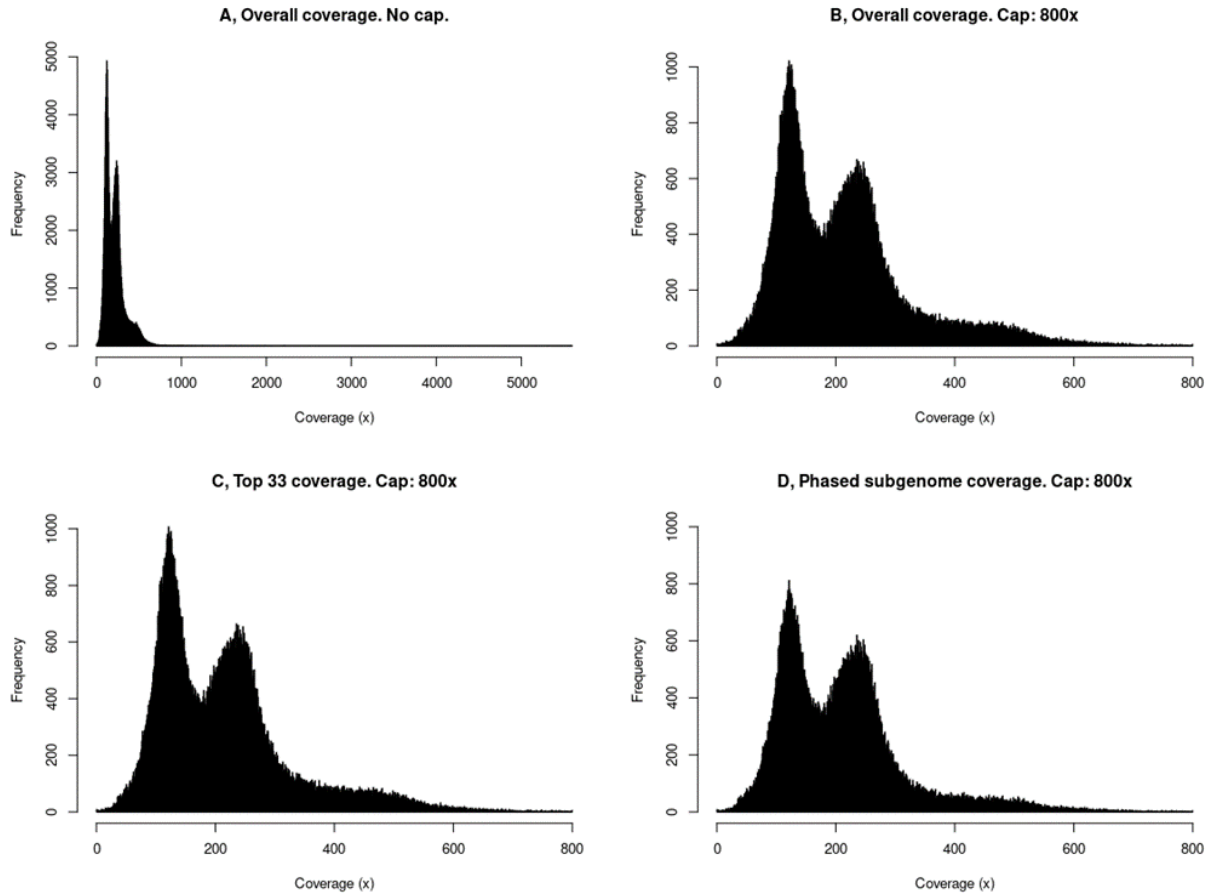
Supplementary Figure 2.3: Chromatin contact map following Hi-C scaffolding and manual curation. Ordered according to length. Generated by *Juicebox*.



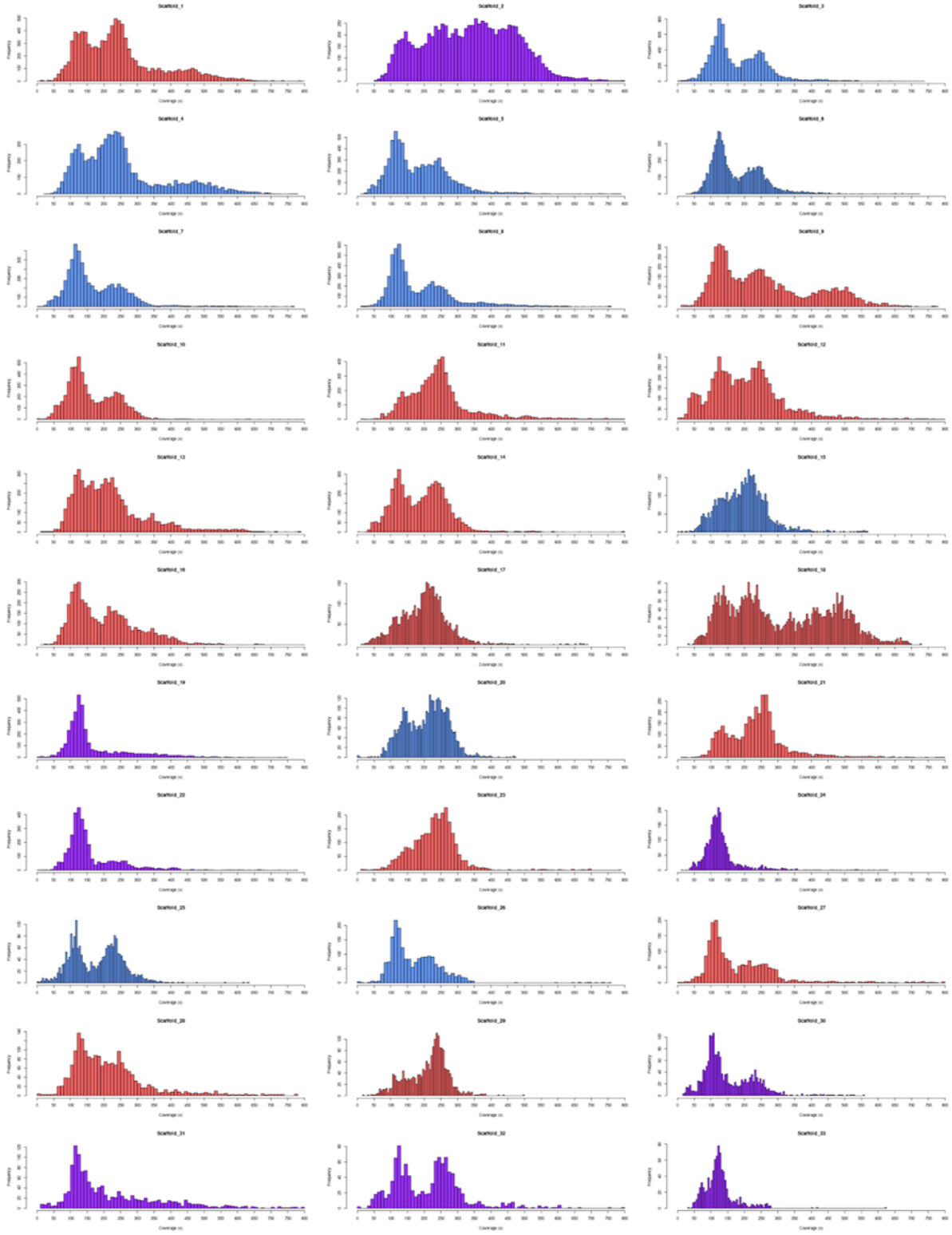
Supplementary Figure 2.4: Total cumulative length over number of scaffolds. Generated using QUAST. X-axis is the number of scaffolds. Y-axis is the cumulative length in megabases. Green dashed line indicates the end of scaffold 33.



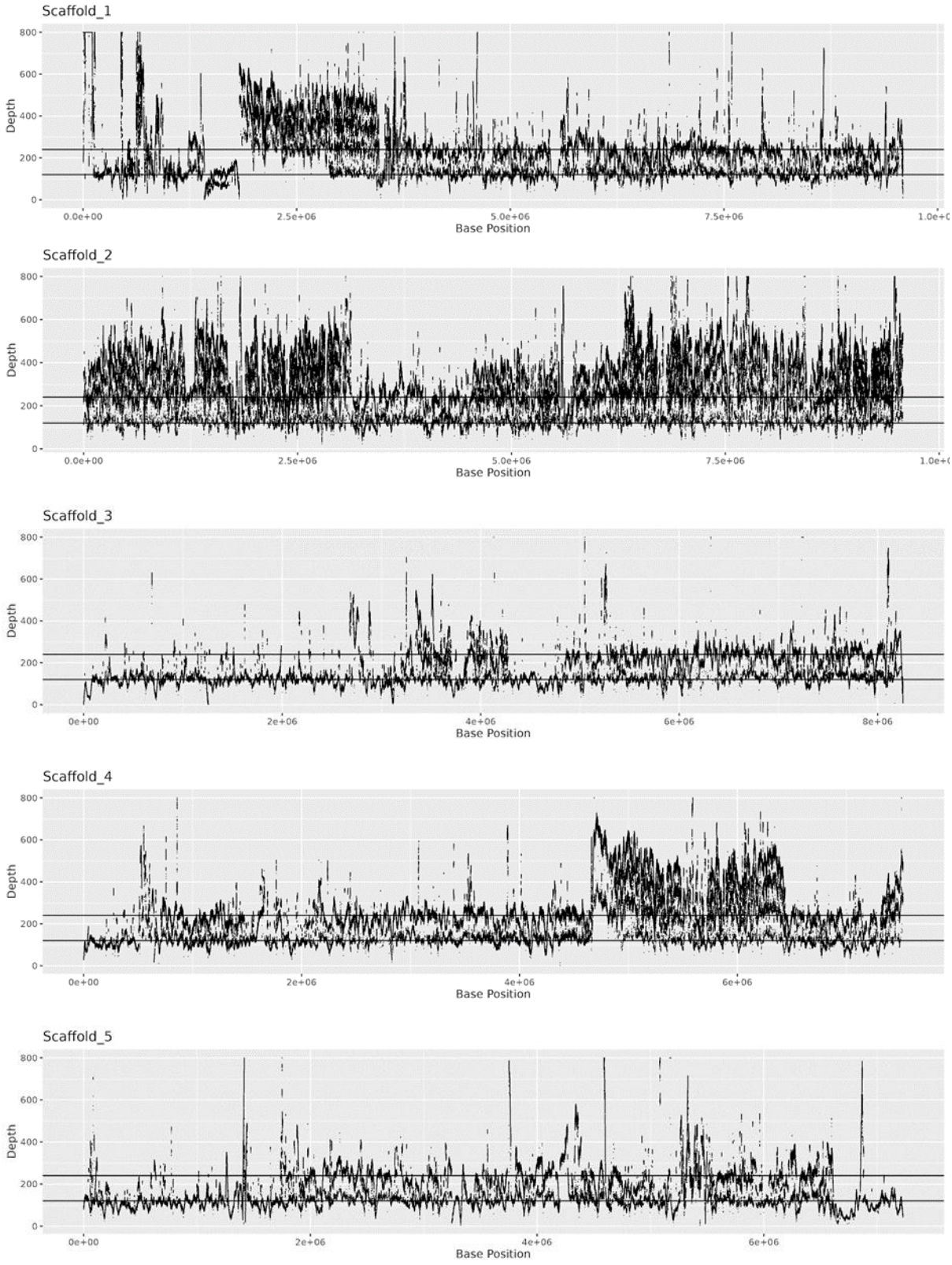
Supplementary Figure 2.5: Blobplot of contamination in our *Meloidogyne javanica* assembly. Generated in *Blobtools.* The x-axis represents the GC proportion, and the y-axis the coverage depth of sequences in the assembly. The size and colour of each blob indicate the relative abundance and taxonomic identity of the contaminating organisms according to the key.



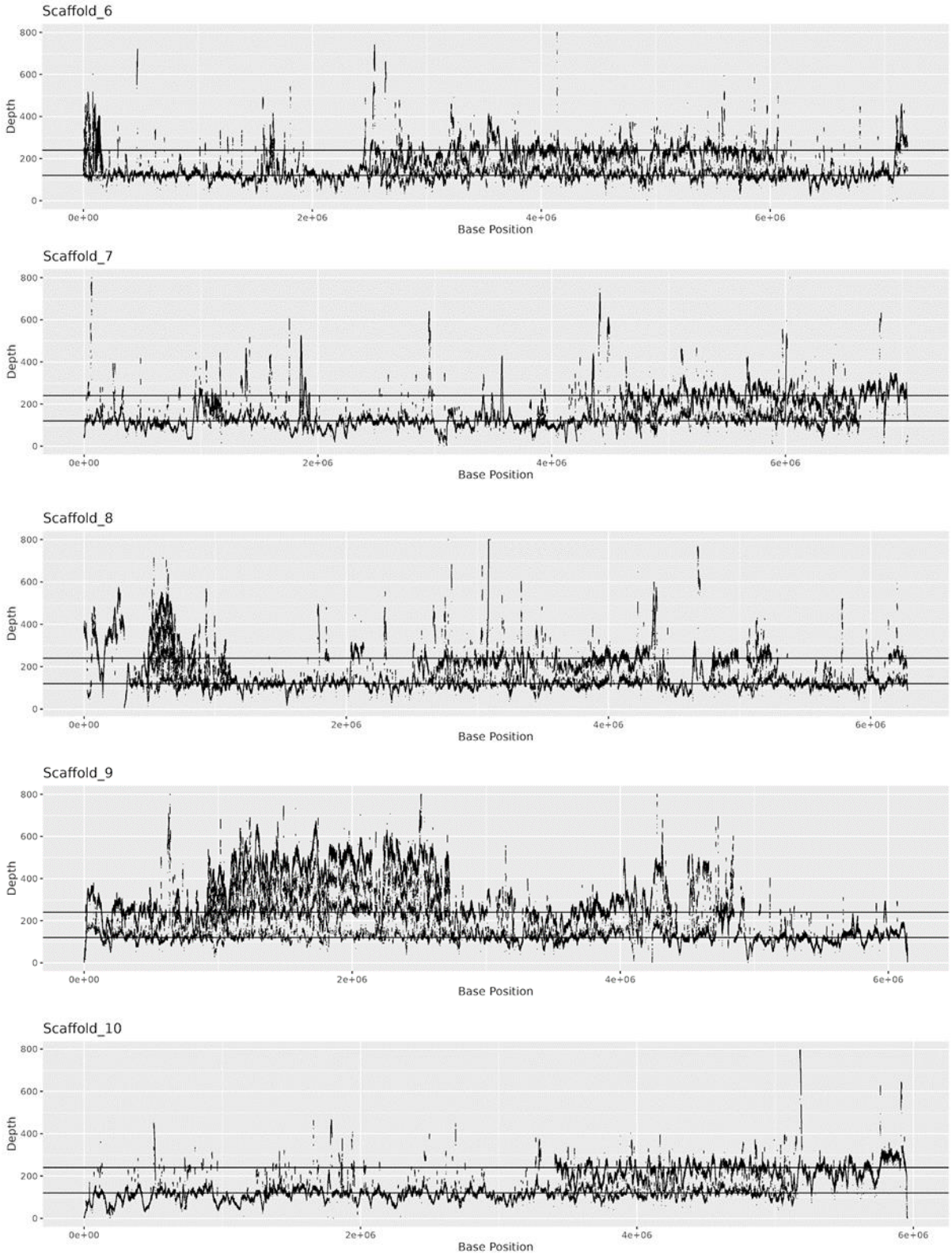
Supplementary Figure 2.6: Assembly-wide distributions of coverage depth. A, Distribution of coverage depth at all bases in the assembly. **B**, Distribution of coverage depth of all bases in the assembly, capped at 800x. **C**, Distribution of coverage of all bases in the longest 33 scaffolds, capped at 800x. **D**, Distribution of coverage of all bases in scaffolds phased to a subgenome, capped at 800x. Breaks = 1000. Bin size = 0.8.



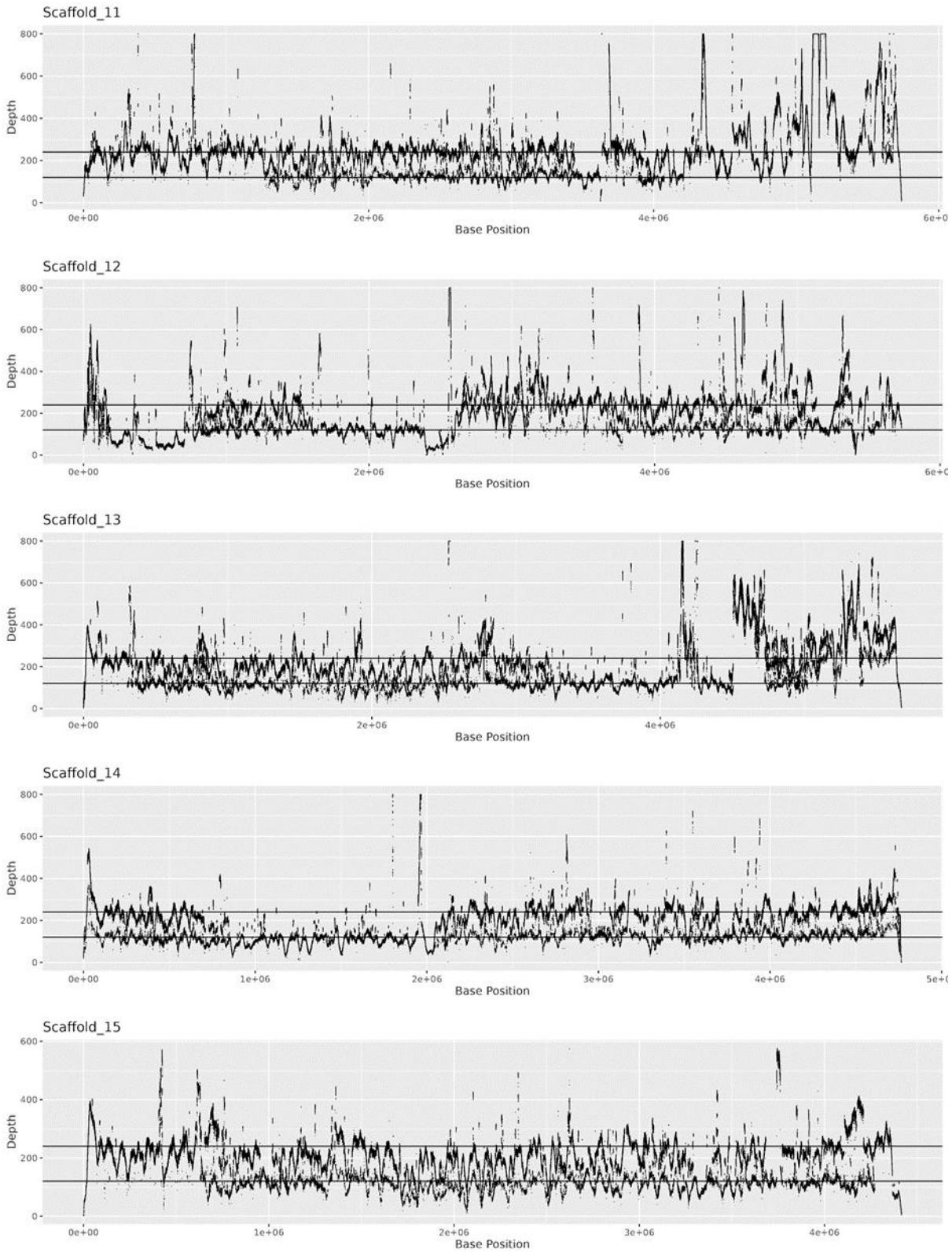
Supplementary Figure 2.7: Scaffold-level distributions of coverage depth frequencies. Subgenome A in blue, subgenome B in red, and unphased scaffolds in purple. X-axis shows coverage depth, Y-axis shows frequency. Breaks = 80. Bin size = 10.



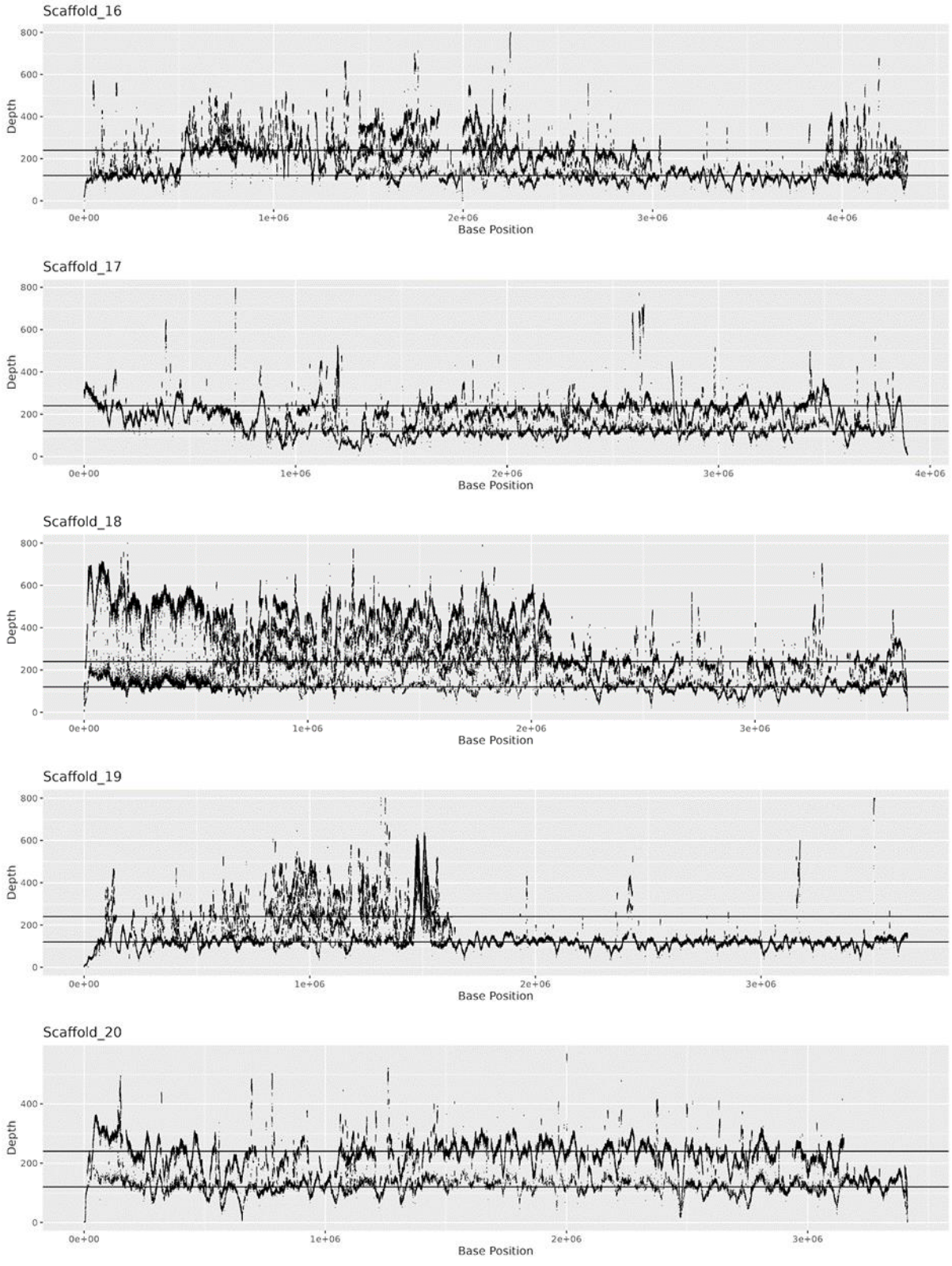
Supplementary Figure 2.8a: Coverage depth of individual bases across scaffolds 1-5. Horizontal lines mark 120x and 240x coverage depth, corresponding to peaks seen in Supplementary Figure 7.



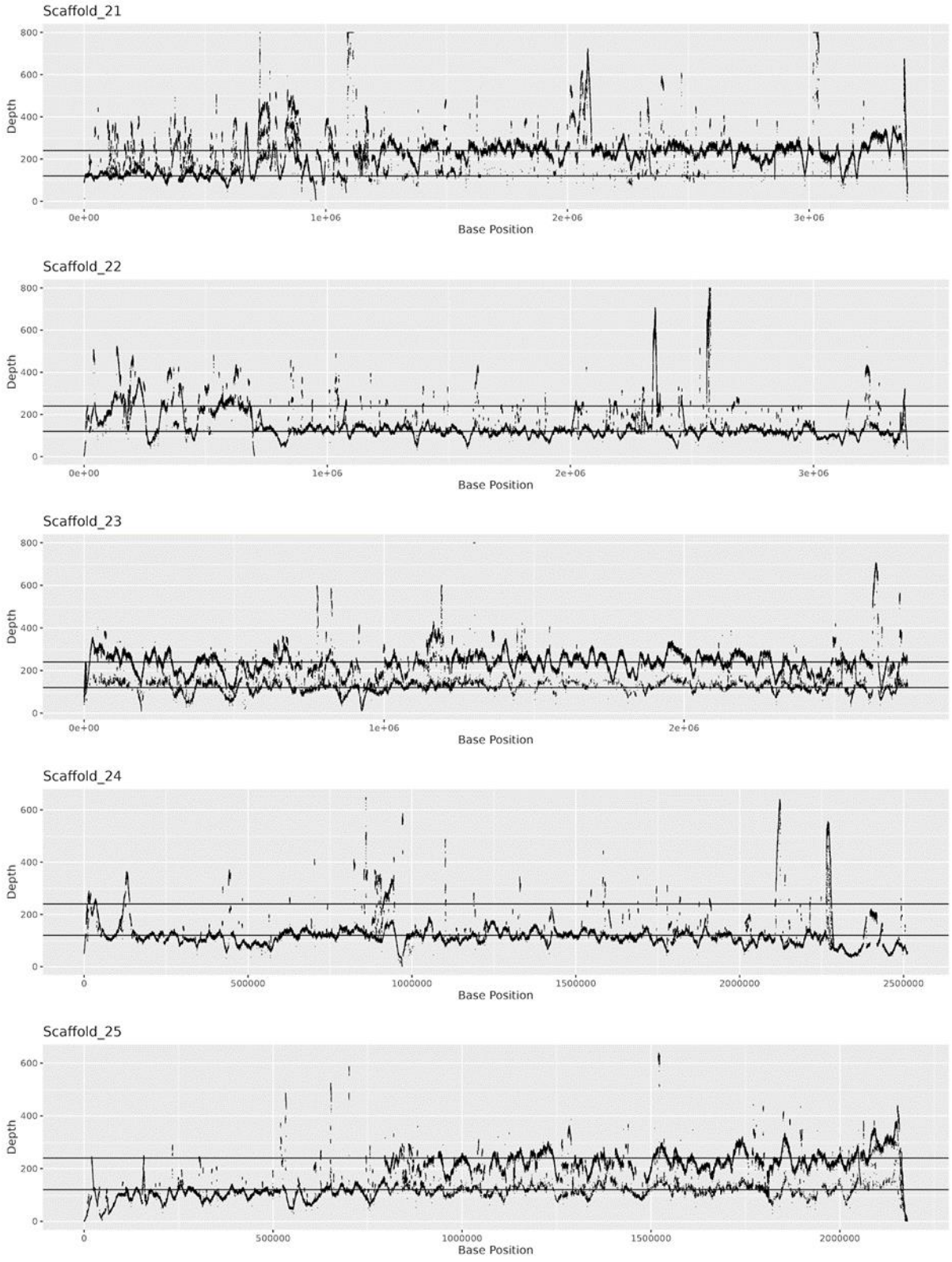
Supplementary Figure 2.8b: Coverage depth of individual bases across scaffolds 6-10. Horizontal lines mark 120x and 240x coverage depth, corresponding to peaks seen in Supplementary Figure 7.



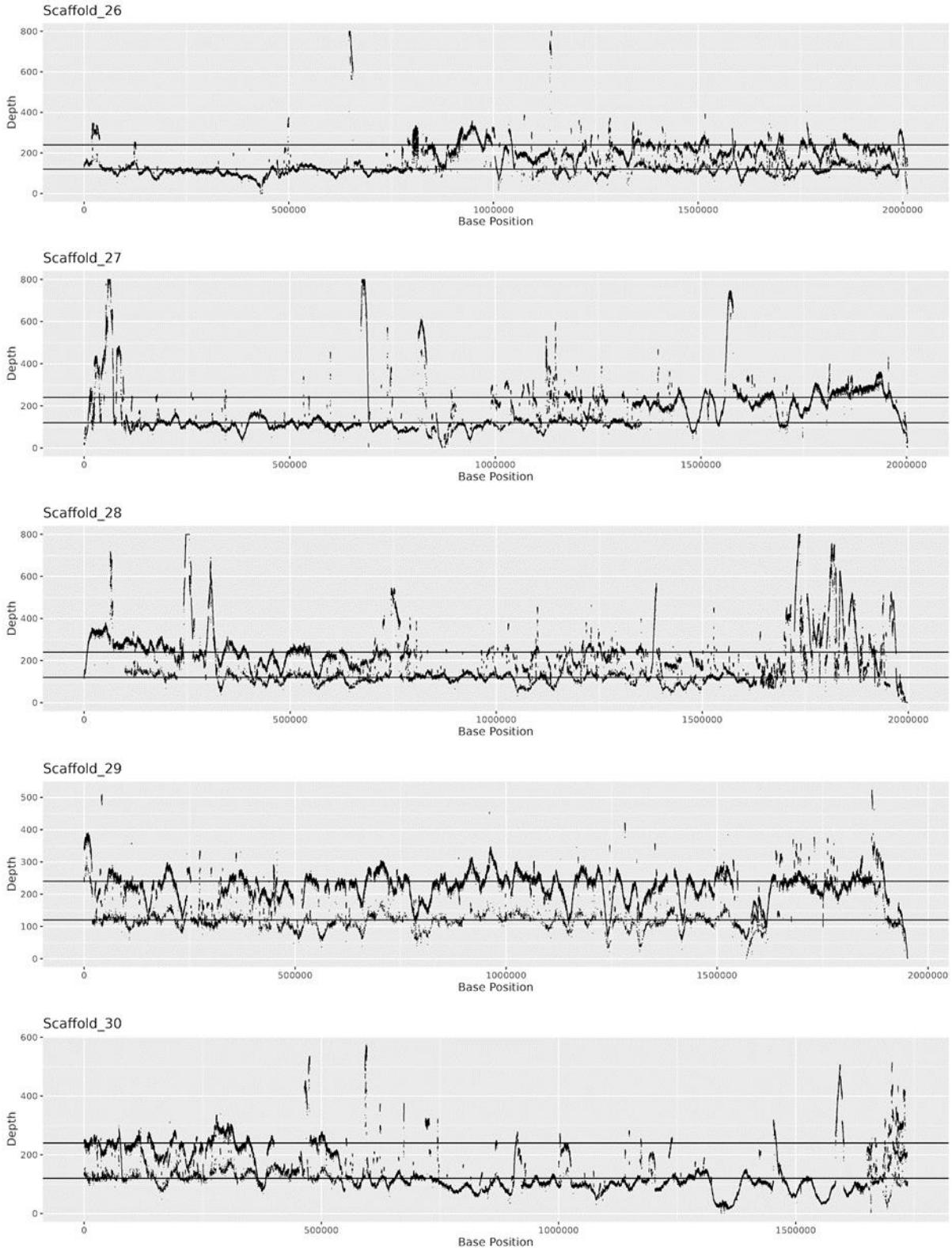
Supplementary Figure 2.8c: Coverage depth of individual bases across scaffolds 11-15. Horizontal lines mark 120x and 240x coverage depth, corresponding to peaks seen in Supplementary Figure 7.



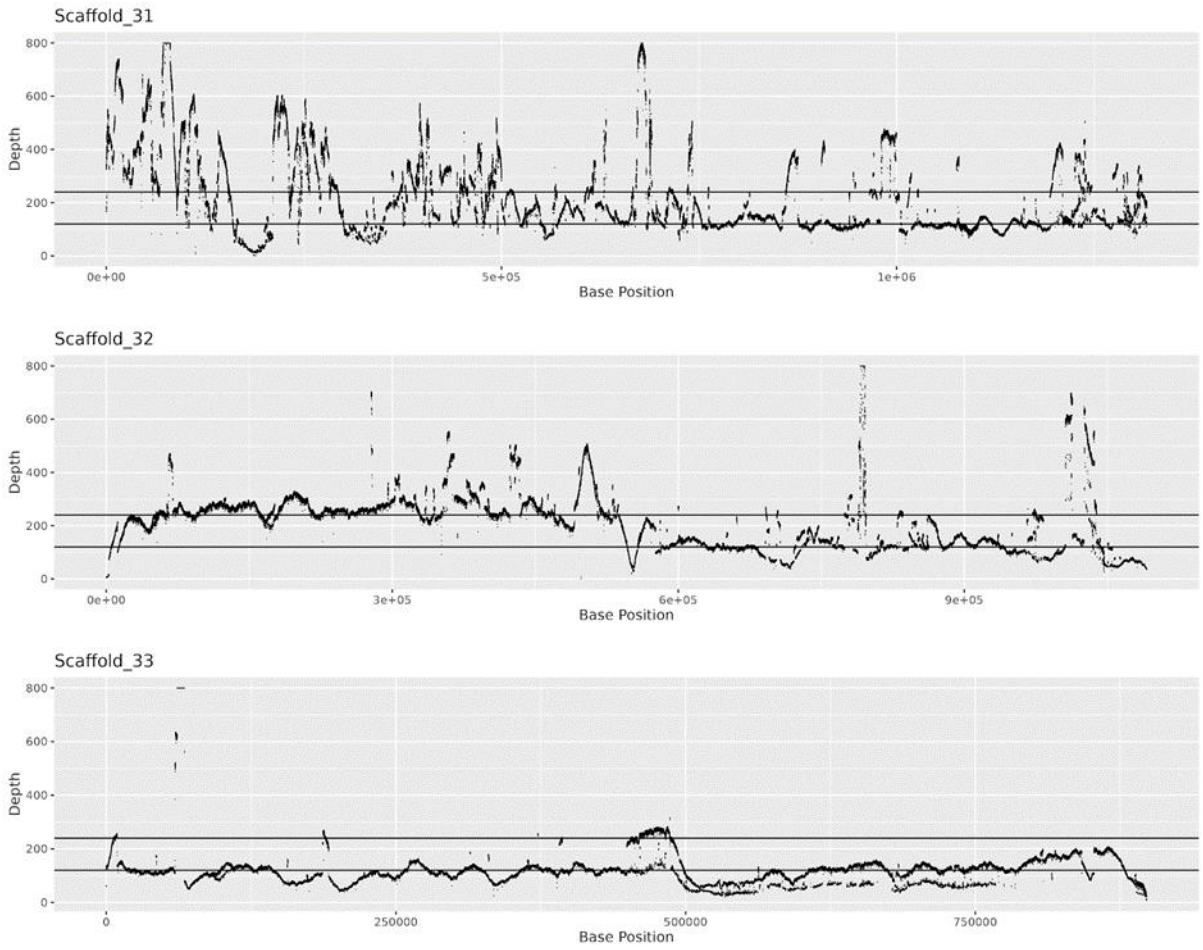
Supplementary Figure 2.8d: Coverage depth of individual bases across scaffolds 16-20. Horizontal lines mark 120x and 240x coverage depth, corresponding to peaks seen in Supplementary Figure 7.



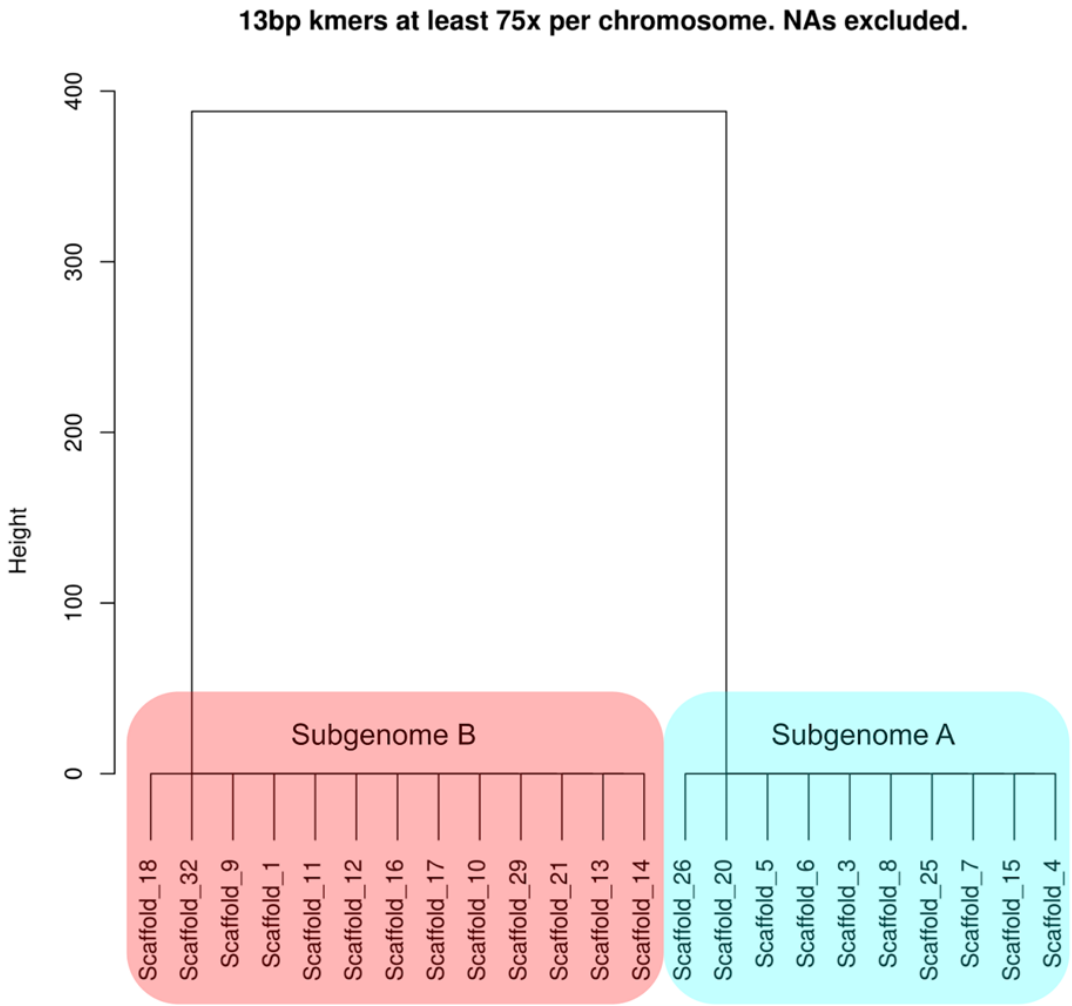
Supplementary Figure 2.8e: Coverage depth of individual bases across scaffolds 21-25. Horizontal lines mark 120x and 240x coverage depth, corresponding to peaks seen in Supplementary Figure 7.



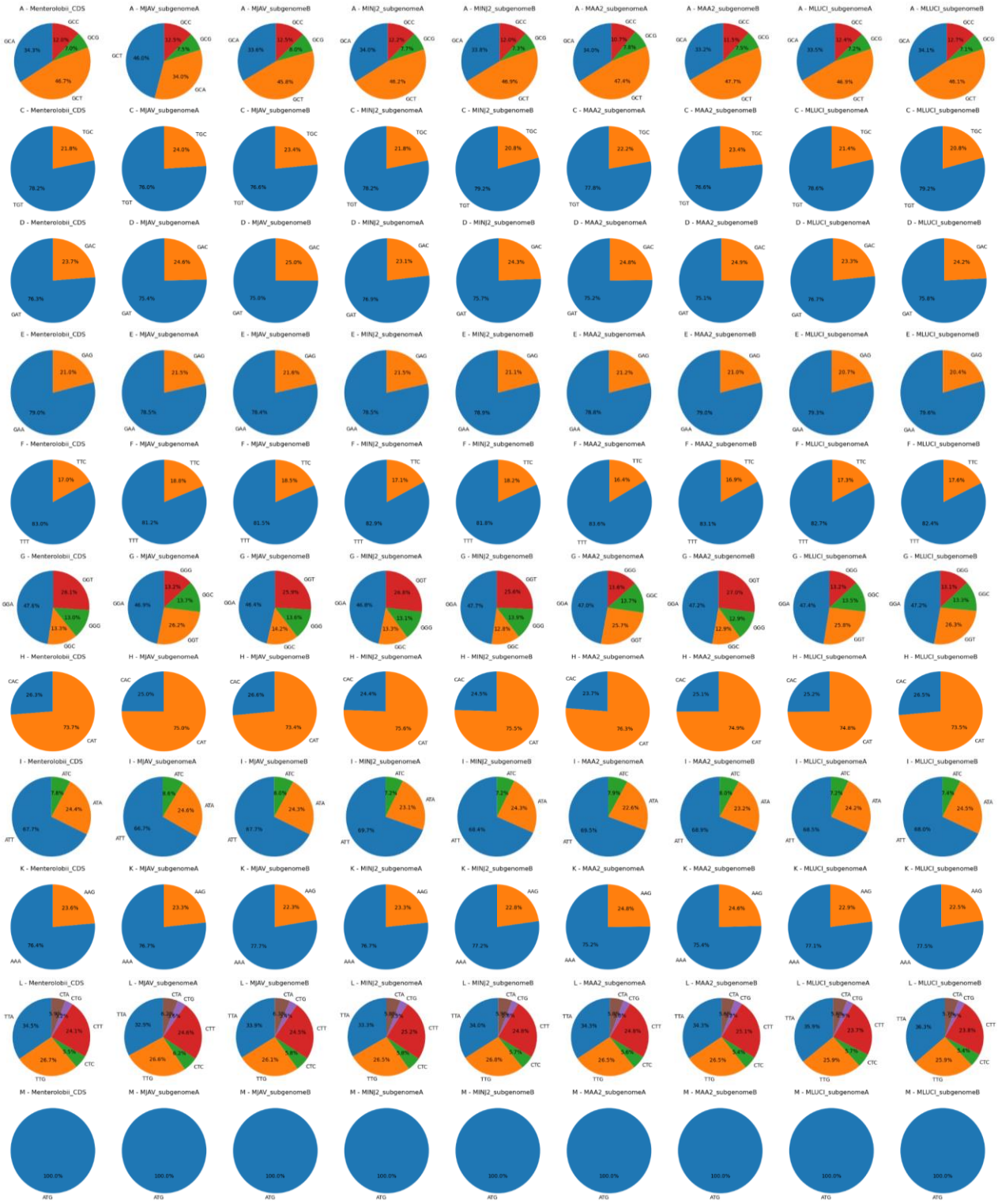
Supplementary Figure 2.8f: Coverage depth of individual bases across scaffolds 26-30. Horizontal lines mark 120x and 240x coverage depth, corresponding to peaks seen in Supplementary Figure 7.



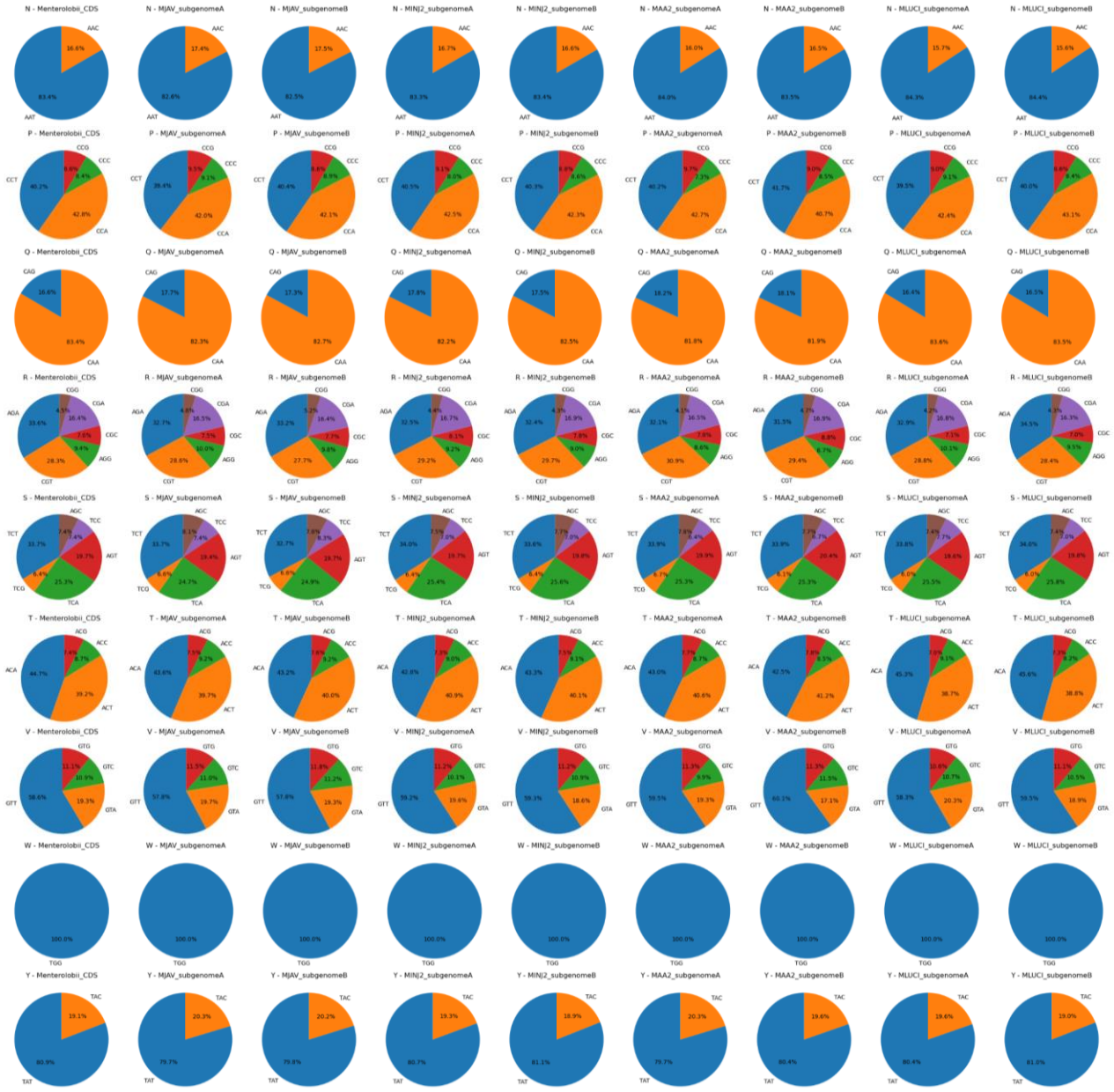
Supplementary Figure 2.8g: Coverage depth of individual bases across scaffolds 31-33. Horizontal lines mark 120x and 240x coverage depth, corresponding to peaks seen in Supplementary Figure 7.



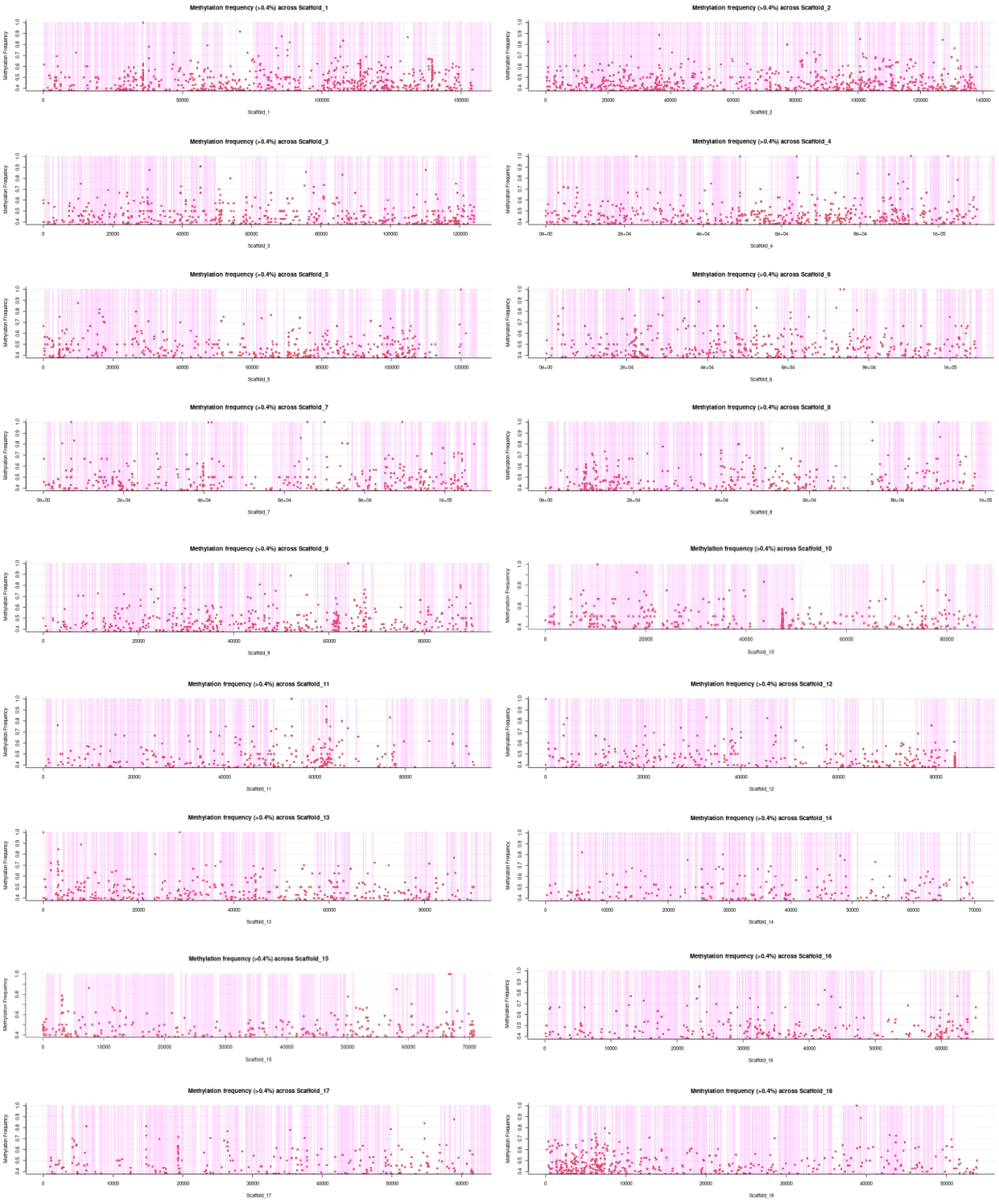
Supplementary Figure 2.9: Subgenome-phased dendrogram of scaffolds. Generated in *R* using hierarchical clustering based on presence of ancestral k-mers. Scaffolds in blue are assigned to subgenome A. Scaffolds in red are assigned to subgenome B.



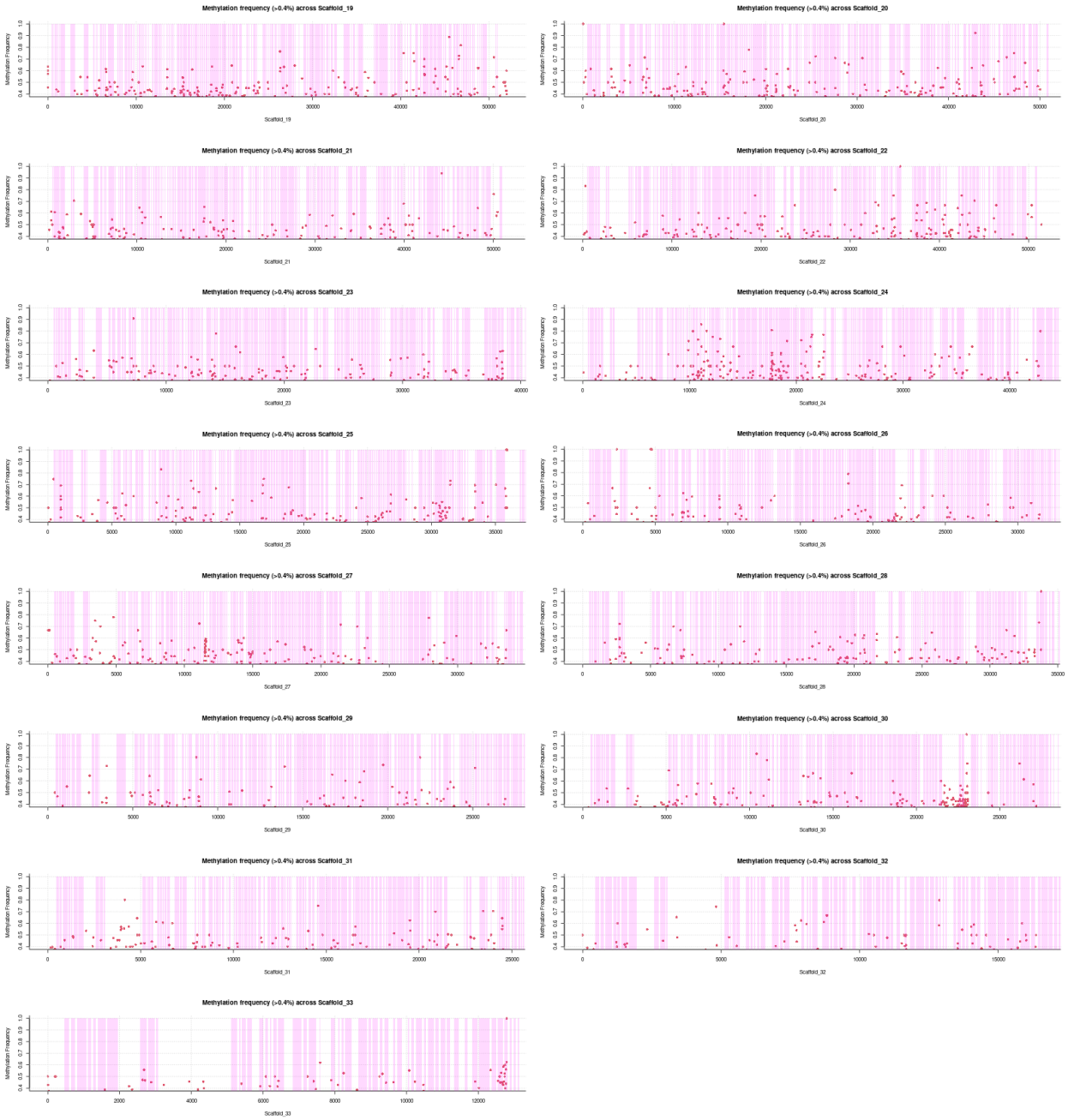
Supplementary Figure 3.1.1: Codon usage frequencies for each subgenome of all species included in these analyses.



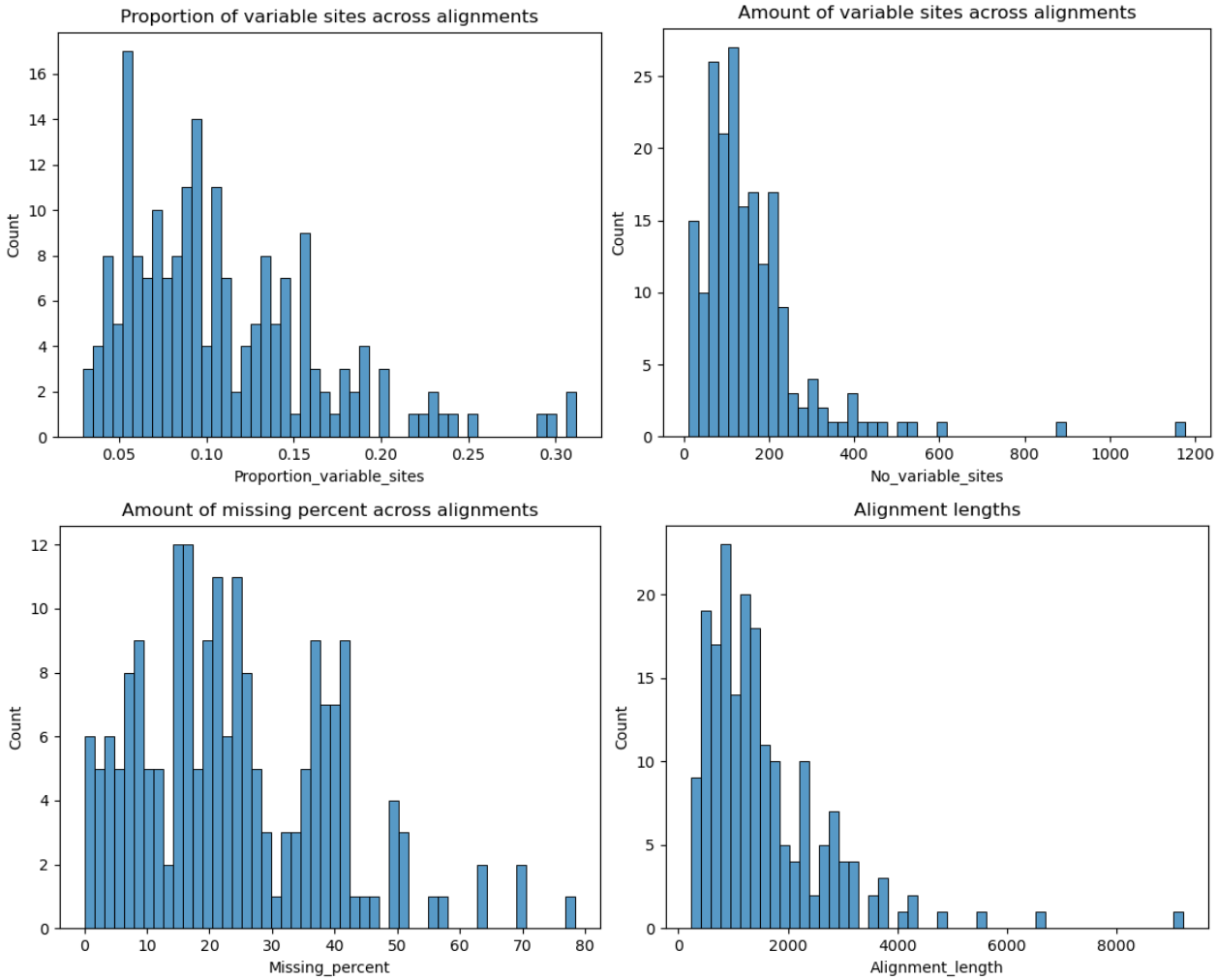
Supplementary Figure 3.1.2: Codon usage frequencies for each subgenome of all species included in these analyses (continued).



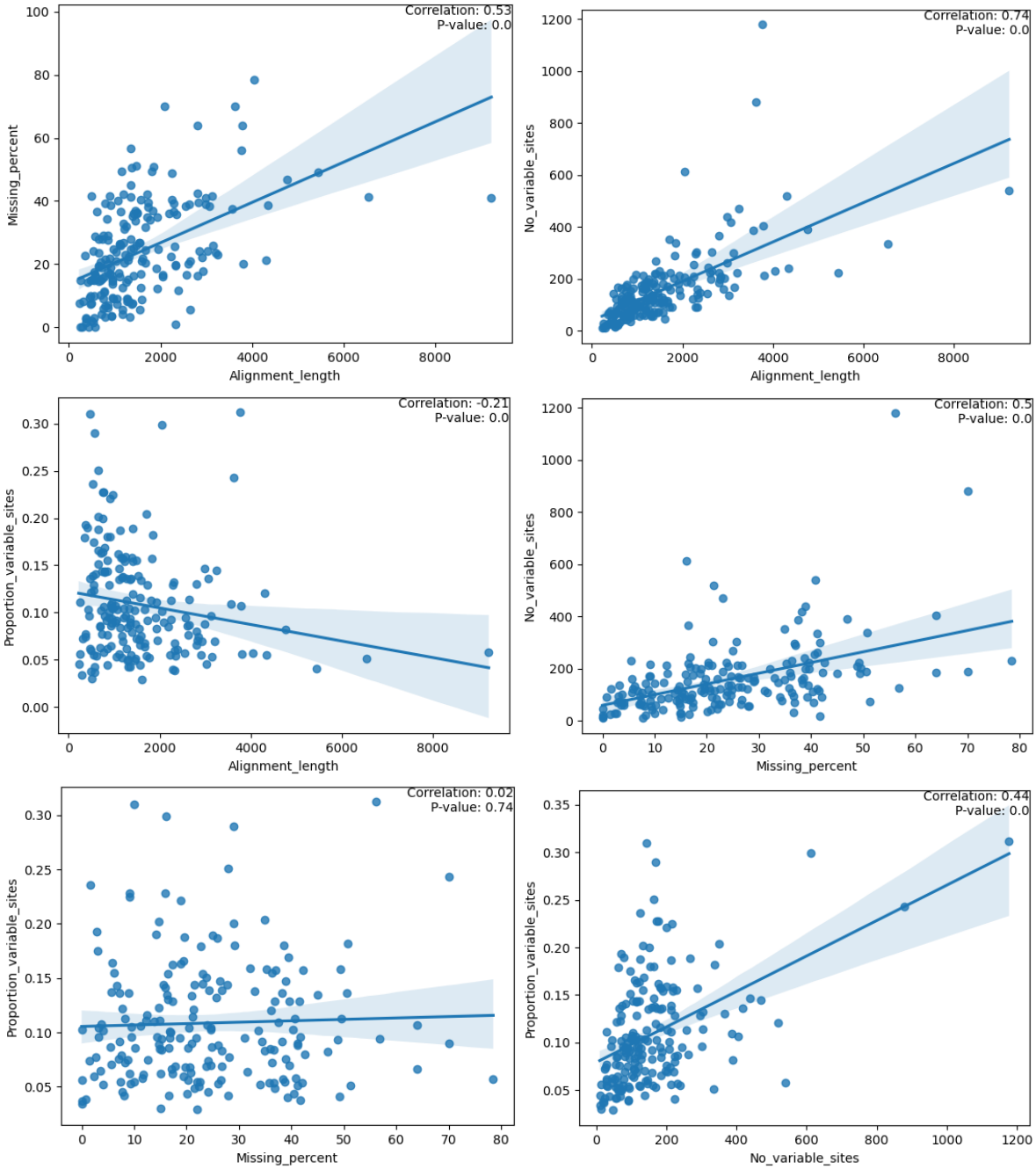
Supplementary Figure 3.2.1: Methylation patterns of scaffolds 1-18. X axis is the base position on the scaffold. Y axis is the frequency of reads covering that base that were methylated. Dots represent bases. Pink bars show CDS positions along the scaffold. Only calls >2 log likelihood and at frequency over 40% are shown.



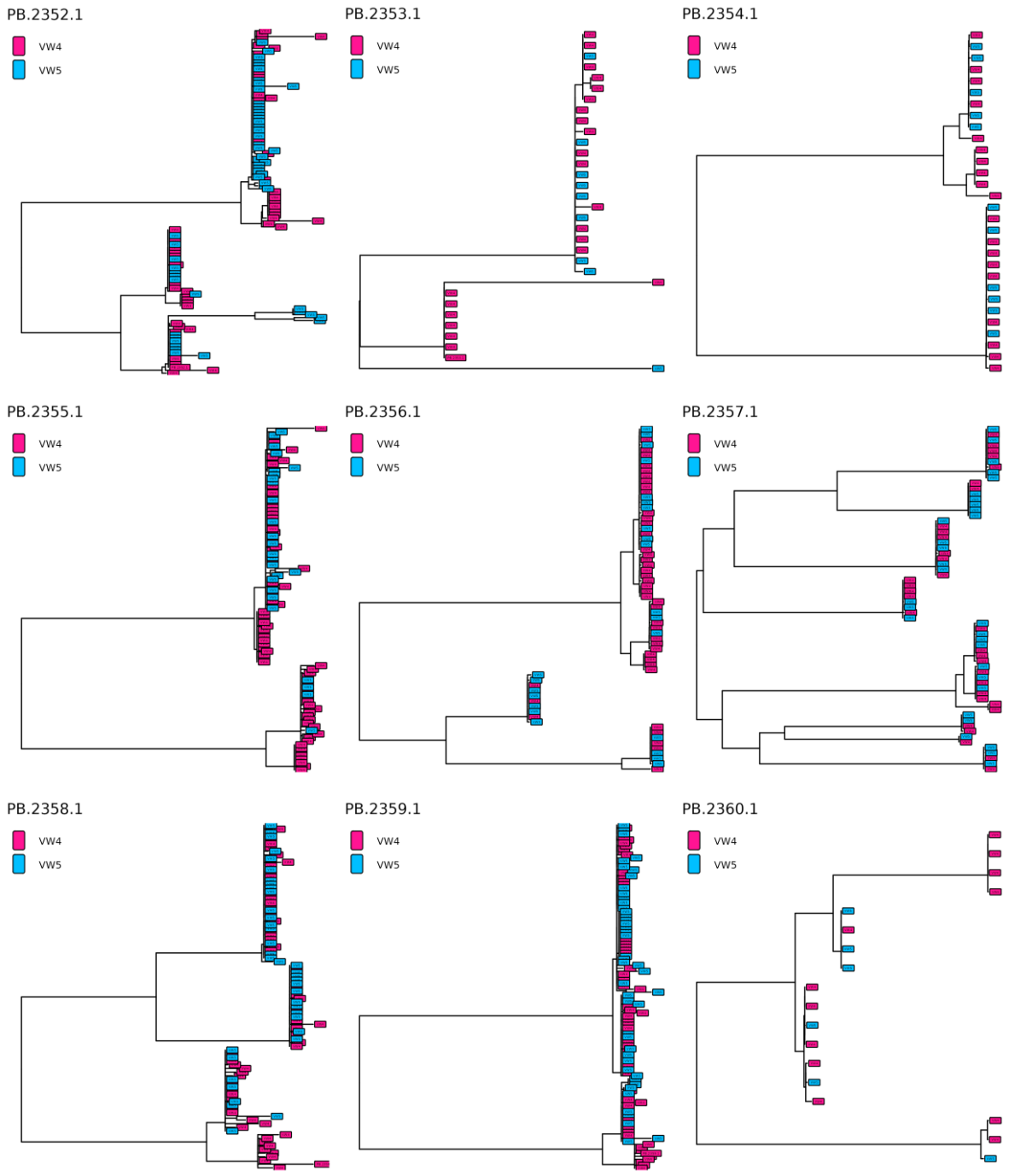
Supplementary Figure 3.2.2: Methylation patterns of scaffolds 19-33. X axis is the base position on the scaffold. Y axis is the frequency of reads covering that base that were methylated. Dots represent bases. Pink bars show CDS positions along the scaffold. Only calls >2 log likelihood and at frequency over 40% are shown.



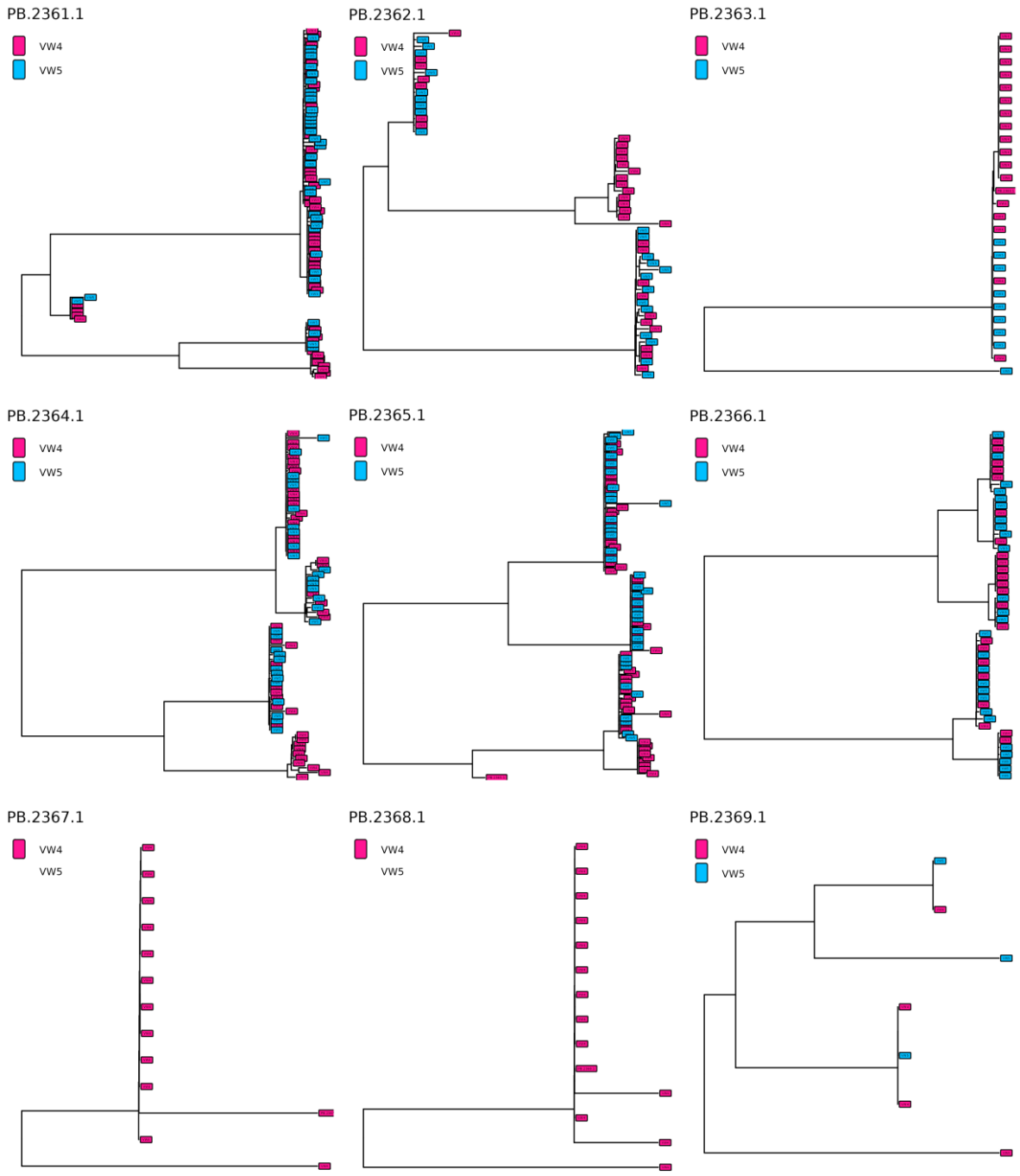
Supplementary Figure 3.3.1: Descriptive statistics of alignments. Top left, proportion of variable sites across alignments. **Top right**, number of variable sites across alignments. **Bottom left**, percent of missing sequence across alignments. **Bottom right**, alignment lengths.



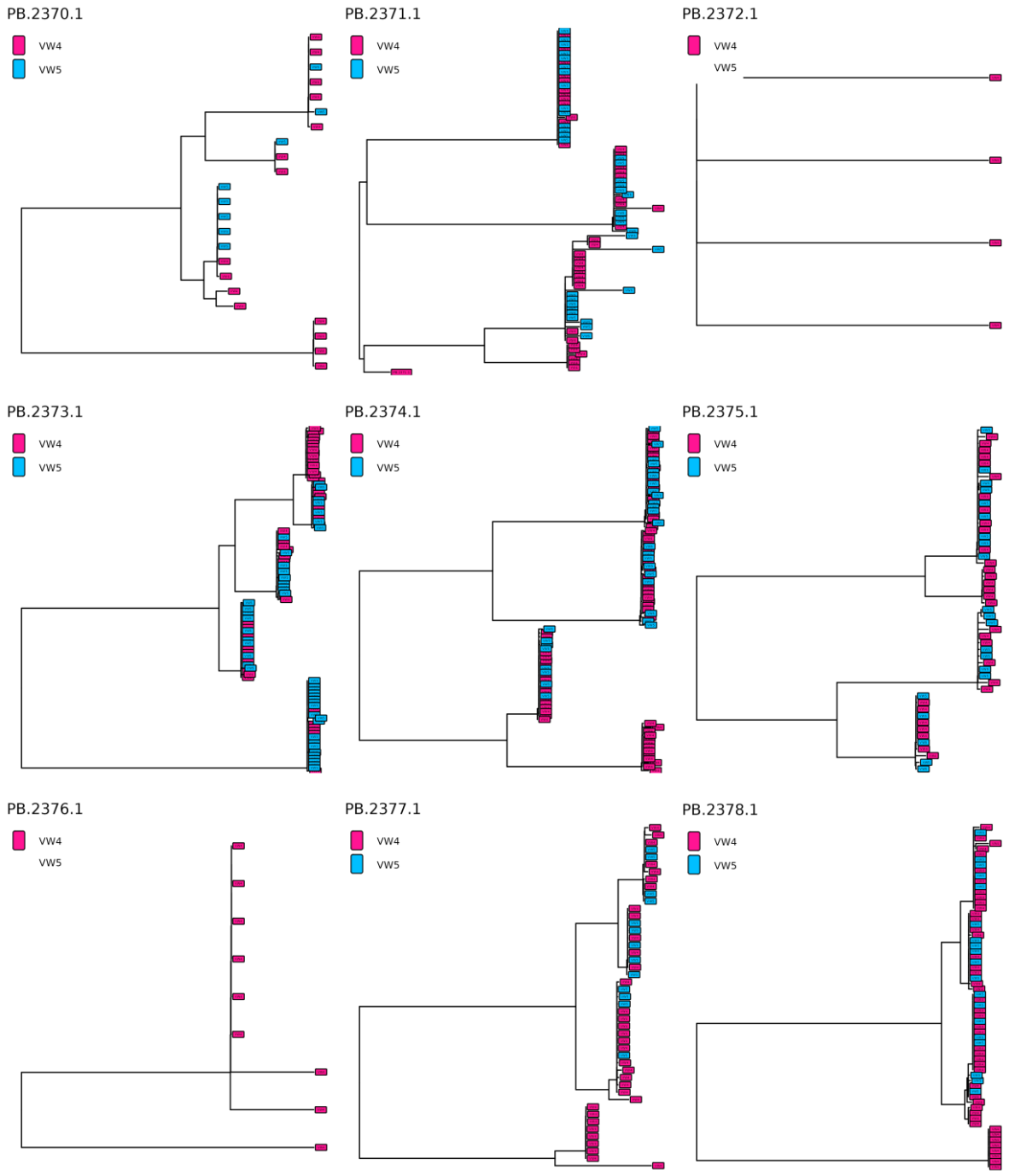
Supplementary Figure 3.3.2: Top left, length of alignment against percent of missing sequence. **Top right**, length of alignment against the number of variable sites. **Middle left**, length of alignment against the proportion of variable sites. **Middle right**, percent of missing sequence against number of variable sites. **Bottom left**, percent of missing sequence against proportion of variable sites. **Bottom right**, number of variable sites against the proportion of variable sites. All comparisons showed positive relationships except length of alignment and proportion of variable sites which showed that proportion of variable sites decreased with an increase in alignment length. All comparisons showed significant relationships (p = 0.0) except the relationship of percent of missing sequence to proportion of variable sites (p = 0.74).



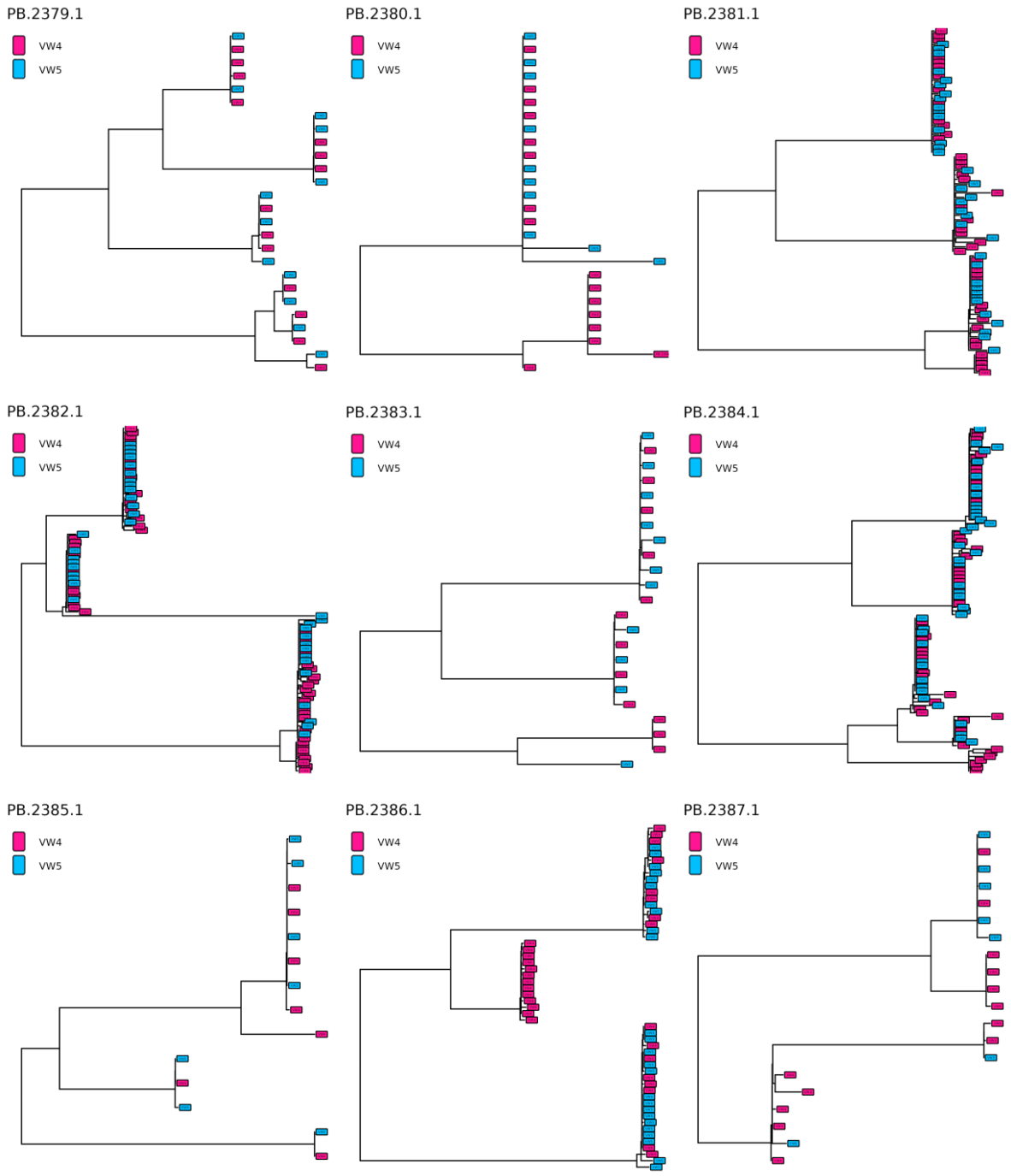
Supplementary Figure 4.1.1: Phylogenetic trees of genes identified within the deleted region of scaffold 15. PB.2354.1 could be indicative of gene conversion/homoeologous recombination. PB.2355.1, PB.2358.1, and PB.2359.1 show the same topology as Gene P and Gene R: four clear clusters, with one cluster containing no representatives from VW5.



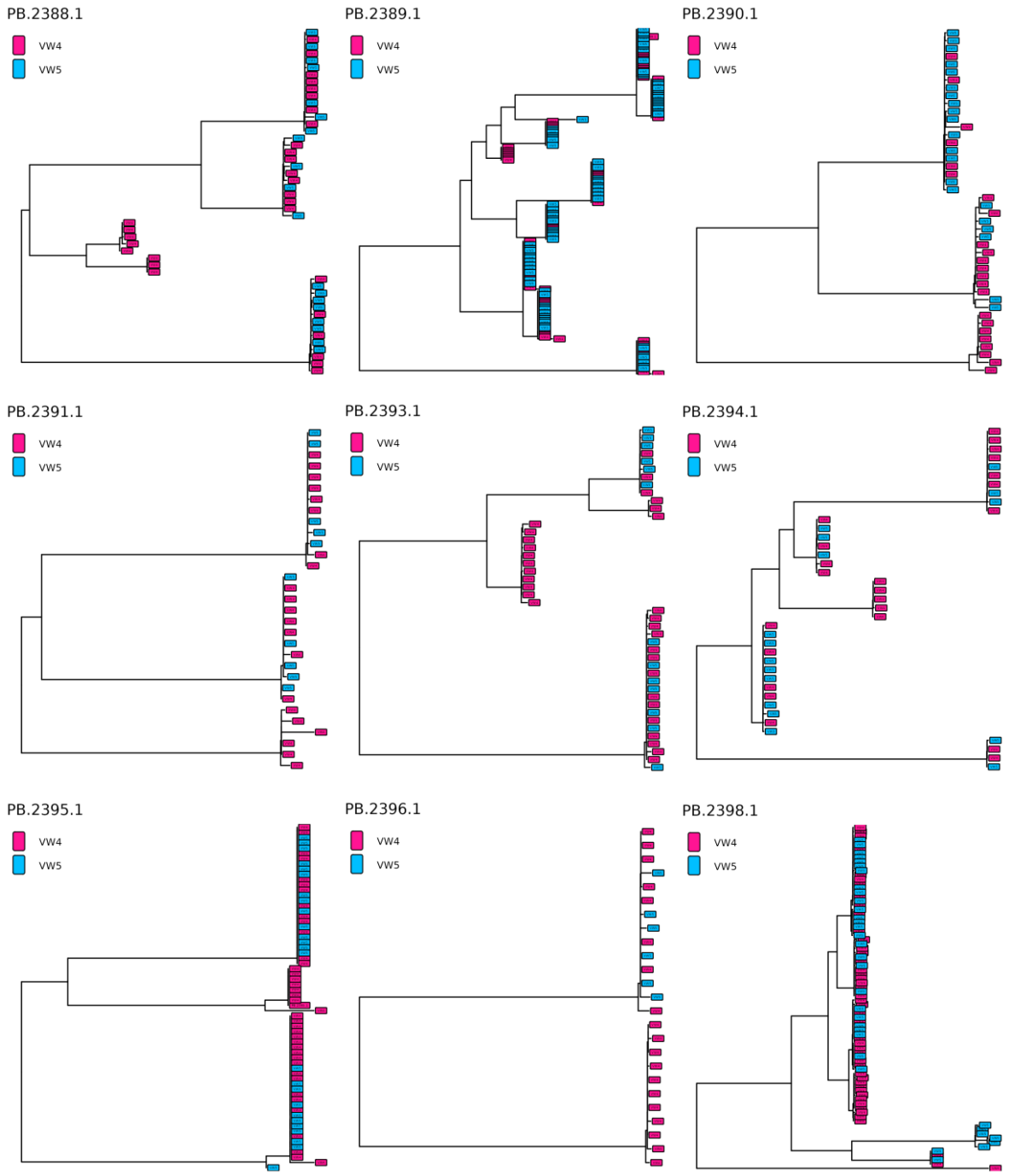
Supplementary Figure 4.1.2: Phylogenetic trees of genes identified within the deleted region of scaffold 15. PB.2361.1, PB.2362.1, PB.2364.1 and PB.2365.1 show the same topology as Gene P and Gene R: four clear clusters, with one cluster containing no representatives from VW5. PB.2367.1 and PB.2368.1 retain no copies from VW5.



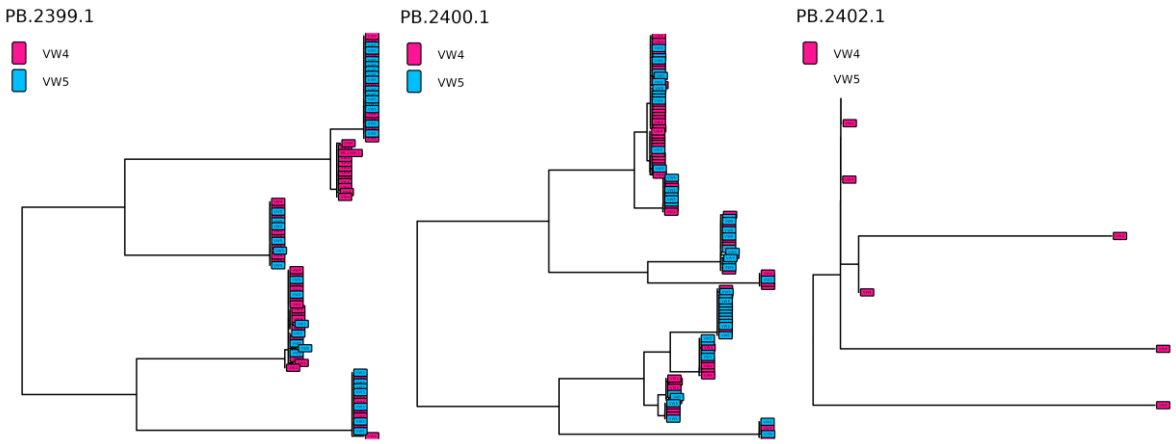
Supplementary Figure 4.1.3: Phylogenetic trees of genes identified within the deleted region of scaffold 15. PB.2374.1, PB.2375.1, and PB.2378.1 show similar topology as Gene P and Gene R: four clear clusters, with one cluster containing no representatives from VW5. PB.2372.1 is likely an erroneous annotation, with the sequence present in only 4 reads.



Supplementary Figure 4.1.4: Phylogenetic trees of genes identified within the deleted region of scaffold 15. PB.2380.1 shows only two clusters, with no copies from VW5 in one cluster. PB.2381.1, PB.2381.1, PB.2384.1 and PB.2386.1 show the same topology as Gene P and Gene R: four clear clusters, with one cluster containing no representatives from VW5.



Supplementary Figure 4.1.5: Phylogenetic trees of genes identified within the deleted region of scaffold 15. PB.2390.1 and PB.2391.1 show only three clusters, with only two containing VW5. This indicates that there are no copies of these genes on one subgenome of VW5, and that the region of that subgenome containing these gene sequences is a single copy in VW4. PB.2393.1 and PB.2395.1 also show only three copies, however the cluster containing only VW4 is not on a lone branch, but instead shares a node with a closely related copy. PB.2396.1 is also suggested to be a cathepsin-like gene or PLCP.



Supplementary Figure 4.1.6: Phylogenetic trees of genes identified within the deleted region of scaffold 15. Only 6 reads containing sequences similar to PB.402.1 were found in VW4, and none in VW5.

7.2 Supplementary Tables

Supplementary Table 2.1: Descriptive statistics of assembly and its contemporaries.

Accession	Species	Year	Scaffolds	Assembly size (Mbp)	Nuclear DNA Content (Mbp)	N50 (Kbp)	CEGMA % (complete: C, partial: P)	BUSCO v5 % (complete: C [single: S, duplicated: D], fragmented: F, missing: M)	GC %
GCA_034785575.1	*M. javanica** (Hull)	2023	69	150.5	NA	5,793	C :93.95 (1.88) P: 95.56	C:69.5 [S:37.3%, D:32.2%], F:13.7%, M:16.8%	30.1
GCA_003693625.1	*M. javanica* (Hull)	2017	34,394	142.6	NA	14.1	C :89.52 (2.71) P: 95.16	C:70.2% [S:30.2, D:40.0%], F:13.3%, M:16.5%	30.2
GCA_900003945.1	*M. javanica* (Avignon)	2017	31,341	235.8	297+- 27	10.4	C: 92.74 (3.68) P :95.56	C:70.2% [S:14.9%, D:55.3%], F:14.1%, M:15.7%	30
GCA_014132215.1	*M. incognita* (Morelos)	2017	12,091	183.5	189 +- 15	38.6	C: 94.76 (2.93) P :96.77	C:71.7% [S:18.8%, D:52.9%], F:11.4%, M:16.9%	29.8
GCA_003693645.1	*M. incognita* (Hull)	2017	33,735	122	NA	16.5	C: 82.66 (2.34) P :89.52	C:61.6% [S:29.4%, D:32.2%], F:16.9%, M:21.5%	30.6
GCA_000172435.1	*M. hapla* (VW9)	2008	1,523	53.6	121 +- 3	83.6	C :93.55 (1.19) P: 95.56	C:66.7[S:65.9%, D:0.8%], F:16.1%, M:17.2%	27.4
GCA_003693605.1	*M. floridensis (SJF1)*	2018	9,134	74.9	NA	13.3	C: 77.42 (1.71) P: 83.87	C:58.4% [S:53.3%, D:5.1%], F:17.3%, M:24.3%	30.2
GCA_902706615.1	*M. luci* (SI-Smartno)	2020	327	209.2	NA	1,712	C: 95.56 (2.92) P :96.77	C:73.7% [S:14.9%, D:58.8%], F:11.0%, M:15.3%	30.2
GCA_903994135.1	*M. enterolobii* (Swiss)	2021	4,437	240	275 +- 19	143	C: 94.76 (3.30) P: 96.77	C:73.7% [S:13.3%, D:60.4%], F:10.6%, M:15.7%	30
GCA_900003985.1	*M. arenaria (Guadeloupe)*	2017	26,196	258.1	304 +- 9	16.5	C :94.76 (3.66) P :95.56	C:70.5% [S:12.9%, D:57.6%], F:13.3%, M:16.2%	30
GCA_003133805.1	*M. arenaria (A2-0)*	2019	2,224	284.05	NA	204.6	C: 94.76 (3.57) P :96.77	C:72.6% [S:17.3%, D:55.3%], F:12.2%, M:15.2%	30
GCA_002778205.2	*M. graminicola (IARI)*	2022	4,304	38.18	NA	20.4	C: 84.27 (1.34) P: 90.73	C:65.1% [S:56.1%, D:9.0%], F:13.3%, M:21.6%	23.05

Supplementary Table 2.2: Summary of longest 33 scaffolds of diploid assembly of *M. javanica*.

Scaffold Number	Length (bp)	Mean coverage (x)	Phase assignment (A, B, or unphased)	Comments
1	9595054	245	B	Some collapse.
2	9577269	337	U	Extensive collapse.
3	8254935	181	A	
4	7520906	251	A	Some collapse.
5	7228418	184	A	
6	7199321	177	A	
7	7044505	169	A	
8	6281301	186	A	
9	6143287	260	B	Some collapse.
10	5957753	164	B	
11	5739182	247	B	
12	5730042	205	B	
13	5673700	219	B	
14	4766193	189	B	
15	4415488	196	A	
16	4345108	204	B	Some collapse.
17	3893252	201	B	
18	3682237	314	B	Some collapse.
19	3649086	168	U	Some collapse.
20	3413746	207	A	
21	3407168	264	B	
22	3386470	162	A	Low coverage, possibly single copy. Phased manually.
23	2744999	231	B	Phased manually.
24	2512086	131	U	Low coverage, possibly single copy.
25	2178917	178	A	
26	2011659	171	A	
27	2002407	183	B	Phased manually.
28	1997981	214	B	Phased manually.
29	1951323	210	B	
30	1735511	151	U	
31	1316575	219	U	
32	1091457	211	B	
33	898067	125	B	Low coverage, possibly single copy. Phased manually.

[1] Here collapse refers to regions of a scaffold with more than 2 copies predicted based on read depth.

Supplementary Table 2.3: Tabular breakdown of repeat annotation from *RepeatMasker*.

	Number of elements	Length occupies (bp)	Percentage of sequence (%)
Retroelements	11491	7443793	4.94
SINEs	0	0	0.00
Penelope	0	0	0.00
LINEs	663	657807	0.44
- CRE/SLACS	0	0	0.00
- L2/CR1/Rex	588	633246	0.42
- R1/LOA/Jockey	0	0	0.00
- R2/R4/NeSL	0	0	0.00
- RTE/bOV-b	0	0	0.00
- L1/CIN4	34	3868	0.00
LTR Elements	10828	6785986	4.51
- BEL/Pao	2888	3074846	2.04
- Ty1/Copia	0	0	0.00
- Gypsy/DIRS1	3129	2254025	1.50
DNA Transposons	7992	5674442	3.77
Hobo-Activator	1547	261731	0.17
Tc1-IS630-Pogo	1847	601027	0.40
En-Spm	0	0	0.00
MuDR-IS905	0	0	0.00
PiggyBac	0	0	0.00
Tourist/Harbinger	0	0	0.00
Other (Mirage, P-element, Transib)	0	0	0.00
Rolling-circles	4285	1796873	1.19
Unclassified	129631	25417075	16.88
Total interspersed repeats	-	38535310	25.60
Small RNA	1092	1514780	1.01
Satellites	526	59215	0.04
Simple repeats	55163	2683708	1.78
Low complexity	24100	1270259	0.84

Supplementary Table 2.4: Full table of results from structural annotation with *MAKER3*.

Metric	Count
Number of genes	22433
Number of mrnas	22433
Number of mrnas with utr both sides	2811
Number of mrnas with at least one utr	12486
Number of cds	22433
Number of exons	227617
Number of five_prime_utrs	10044
Number of three_prime_utrs	5253
Number of exon in cds	224453
Number of exon in five_prime_utr	12537
Number of exon in three_prime_utr	5783
Number of intron in cds	202020
Number of intron in exon	205184
Number of intron in five_prime_utr	2493
Number of intron in three_prime_utr	530
Number of single exon gene	91
Number of single exon mrna	91
mean mrnas per gene	1
mean cds per mrna	1
mean exons per mrna	10.1
mean five_prime_utrs per mrna	0.4
mean three_prime_utrs per mrna	0.2
mean exons per cds	10
mean exons per five_prime_utr	1.2
mean exons per three_prime_utr	1.1
mean introns in cds per mrna	9
mean introns in exons per mrna	9.1
mean introns in five_prime_utrs per mrna	0.1
mean introns in three_prime_utrs per mrna	0
Total gene length	70550747
Total mrna length	70550747
Total cds length	29119128
Total exon length	30170066
Total five_prime_utr length	525976
Total three_prime_utr length	524962
Total intron length per cds	40117619
Total intron length per exon	40585865
Total intron length per five_prime_utr	365056

Total intron length per three_prime_utr	85815
mean gene length	3144
mean mrna length	3144
mean cds length	1298
mean exon length	132
mean five_prime_utr length	52
mean three_prime_utr length	99
mean cds piece length	129
mean five_prime_utr piece length	41
mean three_prime_utr piece length	90
mean intron in cds length	198
mean intron in exon length	197
mean intron in five_prime_utr length	146
mean intron in three_prime_utr length	161
Longest genes	85925
Longest mrnas	85925
Longest cds	58605
Longest exons	36616
Longest five_prime_utrs	1124
Longest three_prime_utrs	2255
Longest cds piece	36616
Longest five_prime_utr piece	871
Longest three_prime_utr piece	2230
Longest intron into cds part	46014
Longest intron into exon part	46014
Longest intron into five_prime_utr part	6288
Longest intron into three_prime_utr part	5958
Shortest genes	24
Shortest mrnas	24
Shortest cds	6
Shortest exons	2
Shortest five_prime_utrs	1
Shortest three_prime_utrs	1
Shortest cds piece	1
Shortest five_prime_utr piece	1
Shortest three_prime_utr piece	1
Shortest intron into cds part	5
Shortest intron into exon part	5
Shortest intron into five_prime_utr part	5
Shortest intron into three_prime_utr part	5

Supplementary Table 2.5: List of pairs that share CDS and amount of shared links.

Reference Scaffold	Target Scaffold	Shared CDS
2	13	46
3	11	284
4	14	307
5	12	182
6	19	216
6	1	206
7	10	206
8	16	273
8	33	162
8	29	58
8	28	49
9	5	102
13	15	351
17	25	112
18	26	63
20	23	214
20	32	120
24	30	20
25	27	64

Supplementary Table 2.6: Validation of homoeologous pairings.

Scaffold number	Orthology	MASH	BUSCO	Consensus homoeologous counterpart
1	6	6	6	6
2	13	-	52	13
3	11	11	7	11
4	14	14	6	14
5	12	12	9	12
6	21	21	-	21
7	10	10	-	10
8	16	16	-	16
9	5	-	7	5
10	7	7	3	7
11	3	3	5	3
12	5	5	15	5
13	15	15	45	15
14	4	4	-	4
15	13	13	8	13
16	8	8	25	8
17	25	25	26	25
18	26	-	21	26
19	6	6	30	6
20	23	23	6	23
21	6	6	-	6
22	10	10	-	10
23	20	20	-	20
24	30	-	-	30
25	17	27	-	17
26	18	-	49	18
27	25	-	-	25
28	8	-	8	8
29	8	-	-	8
30	24	-	-	24
31	6	-	-	6
32	20	-	-	20
33	8	-	-	8

Supplementary Table 2.7: Descriptive statistics of each phased subgenome.

	Subgenome A	Subgenome B
Number of sequences	11	17
Length	58,935,666	69,619,210
N50	7,199,321	5,673,700
GC%	29.8	30.27
Largest contig	8,254,935	9,595,054

Supplementary Table 3.1: Primary codon per residue.

	M. javanica		*M. incognita*		*M. arenaria*		*M. luci*		*M. enterolobii*
Amino acid	**A**	**B**	**A**	**B**	**A**	**B**	**A**	**B**	**NA**
Alanine (A)	GCT	GCT	GCT	GCT	GCT	GCT	GCT	GCT	GCT
Arginine (R)	AGA	AGA	AGA	AGA	AGA	AGA	AGA	AGA	AGA
Asparagine (N)	AAT	AAT	AAT	AAT	AAT	AAT	AAT	AAT	AAT
Aspartic Acid (D)	GAT	GAT	GAT	GAT	GAT	GAT	GAT	GAT	GAT
Cysteine (C)	TGT	TGT	TGT	TGT	TGT	TGT	TGT	TGT	TGT
Glutamine (Q)	CAA	CAA	CAA	CAA	CAA	CAA	CAA	CAA	CAA
Glutamic Acid (E)	GAA	GAA	GAA	GAA	GAA	GAA	GAA	GAA	GAA
Glycine (G)	GGA	GGA	GGA	GGA	GGA	GGA	GGA	GGA	GGA
Histidine (H)	CAT	CAT	CAT	CAT	CAT	CAT	CAT	CAT	CAT
Isoleucine (I)	ATT	ATT	ATT	ATT	ATT	ATT	ATT	ATT	ATT
Leucine (L)	TTA	TTA	TTA	TTA	TTA	TTA	TTA	TTA	TTA
Lysine (K)	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA
Methionine (M)	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATG	
Phenylalanine (F)	TTT	TTT	TTT	TTT	TTT	TTT	TTT	TTT	TTT
Proline (P)	CCA	CCA	CCA	CCA	CCA	CCA	CCA	CCA	CCA
Serine (S)	TCT	TCT	TCT	TCT	TCT	TCT	TCT	TCT	TCT
Threonine (T)	ACA	ACA	ACA	ACA	ACA	ACA	ACA	ACA	ACA
Tryptophan (W)	TGG	TGG	TGG	TGG	TGG	TGG	TGG	TGG	TGG
Tyrosine (Y)	TAT	TAT	TAT	TAT	TAT	TAT	TAT	TAT	TAT
Valine (V)	GTT	GTT	GTT	GTT	GTT	GTT	GTT	GTT	GTT

Supplementary Table 3.2.1: Average number of genes per-species in orthogroup

Average number of genes per-species in orthogroup	Number of orthogroups	Percentage of orthogroups	Number of genes	Percentage of genes
<1	11,974	62.4	48,928	34.1
1	6,366	33.2	74,985	52.2
2	681	3.6	14,096	9.8
3	112	0.6	3,334	2.3
4	20	0.1	795	0.6
5	12	0.1	567	0.4
6	4	0	238	0.2
7	2	0	130	0.1
8	4	0	303	0.2
9	0	0	0	0
10	0	0	0	0
11-15	2	0	259	0.2
16+	0	0	0	0

Supplementary Table 3.2.2: Number of species in orthogroups

Number of species in orthogroup	Number of orthogroups
1	2,223
2	4,698
3	2,761
4	1,867
5	2,323
6	1,692
7	1,397
8	1,374
9	842

Supplementary Table 3.3.1: Subgenome specific orthogroup statistics.

Species	*M. arenaria*		*M. incognita*		*M. javanica*		*M. luci*		*M. enterolobii*
Subgenome	A	B	A	B	A	B	A	B	NA
Number of genes	5178	10997	6031	11270	9412	11076	17316	22319	59773
Number of genes in orthogroups	4997	10448	5832	10820	8993	10531	15313	19894	56807
Number of unassigned genes	181	549	199	450	419	545	2003	2425	2966
Percentage of genes in orthogroups	96.5	95	96.7	96	95.5	95.1	88.4	89.1	95
Percentage of unassigned genes	3.5	5	3.3	4	4.5	4.9	11.6	10.9	5
Number of orthogroups containing species	3675	6861	4587	7733	7133	8289	10218	12079	16911
Percentage of orthogroups containing species	19.2	35.8	23.9	40.3	37.2	43.2	53.3	63	88.2
Number of species-specific orthogroups	20	62	18	60	113	60	139	205	1546
Number of genes in species-specific orthogroups	41	129	39	124	289	146	307	474	5556
Percentage of genes in species-specific orthogroups	0.8	1.2	0.6	1.1	3.1	1.3	1.8	2.1	9.3

Supplementary Table 3.3.2: Subgenome specific orthogroup statistics - number of genes per orthogroup (orthogroup counts).

	Number of orthogroups for OTU								
	M. arenaria		***M. incognita***		***M. javanica***		***M. luci***		***M. enterolobii***
Number of genes per-species in orthogroup	A	B	A	B	A	B	A	B	NA
0	15502	12316	14590	11444	12044	10888	8959	7098	2266
1	2704	4325	3589	5403	5796	6535	6706	6884	3420
2	745	1815	809	1796	1018	1469	2527	3566	3586
3	147	495	148	379	211	168	641	1046	4317
4	52	158	28	110	66	77	226	367	2527
5	14	43	10	27	18	21	64	121	1151
6	11	18	2	14	17	9	25	51	695
7	0	4	1	3	1	4	17	21	388
8	1	2	0	1	2	1	3	11	245
9	0	1	0	0	2	2	1	3	135
10	1	0	0	0	0	2	2	4	94
11-15	0	0	0	0	1	1	4	5	205
16-20	0	0	0	0	1	0	2	0	61
21-50	0	0	0	0	0	0	0	0	80
51-100	0	0	0	0	0	0	0	0	5
101-150	0	0	0	0	0	0	0	0	2
151+	0	0	0	0	0	0	0	0	0

Supplementary Table 3.3.3: Subgenome specific orthogroup statistics - number of genes per orthogroup (orthogroup percentages).

	Percentage of orthogroups (%)								
	M. arenaria		***M. incognita***		***M. javanica***		***M. luci***		***M. enterolobii***
Number of genes per-species in orthogroup	A	B	A	B	A	B	A	B	NA
0	80.8	64.2	76.1	59.7	62.8	56.8	46.7	37	11.8
1	14.1	22.6	18.7	28.2	30.2	34.1	35	35.9	17.8
2	3.9	9.5	4.2	9.4	5.3	7.7	13.2	18.6	18.7
3	0.8	2.6	0.8	2	1.1	0.9	3.3	5.5	22.5
4	0.3	0.8	0.1	0.6	0.3	0.4	1.2	1.9	13.2
5	0.1	0.2	0.1	0.1	0.1	0.1	0.3	0.6	6
6	0.1	0.1	0	0.1	0.1	0	0.1	0.3	3.6
7	0	0	0	0	0	0	0.1	0.1	2
8	0	0	0	0	0	0	0	0.1	1.3
9	0	0	0	0	0	0	0	0	0.7
10	0	0	0	0	0	0	0	0	0.5
11-15	0	0	0	0	0	0	0	0	1.1
16-20	0	0	0	0	0	0	0	0	0.3
21-50	0	0	0	0	0	0	0	0	0.4
51+	0	0	0	0	0	0	0	0	0

Supplementary Table 3.3.4: Subgenome specific orthogroup statistics - number of genes per orthogroup (gene counts).

	Number of genes								
	M. arenaria		***M. incognita***		***M. javanica***		***M. luci***		***M. enterolobii***
Number of genes per-species in orthogroup	A	B	A	B	A	B	A	B	NA
0	0	0	0	0	0	0	0	0	0
1	2704	4325	3589	5403	5796	6535	6706	6884	3420
2	1490	3630	1618	3592	2036	2938	5054	7132	7172
3	441	1485	444	1137	633	504	1923	3138	12951
4	208	632	112	440	264	308	904	1468	10108
5	70	215	50	135	90	105	320	605	5755
6	66	108	12	84	102	54	150	306	4170
7	0	28	7	21	7	28	119	147	2716
8	8	16	0	8	16	8	24	88	1960
9	0	9	0	0	18	18	9	27	1215
10	10	0	0	0	0	20	20	40	940
11-15	0	0	0	0	11	13	49	59	2561
16-20	0	0	0	0	20	0	35	0	1083
21-50	0	0	0	0	0	0	0	0	2191
51-100	0	0	0	0	0	0	0	0	310
101-150	0	0	0	0	0	0	0	0	255
151+	0	0	0	0	0	0	0	0	0

Supplementary Table 3.3.5: Subgenome specific orthogroup statistics - number of genes per orthogroup (gene percentages).

	Percentage of genes (%)								
	M. arenaria		***M. incognita***		***M. javanica***		***M. luci***		***M. enterolobii***
Number of genes per-species in orthogroup	A	B	A	B	A	B	A	B	NA
0	0	0	0	0	0	0	0	0	0
1	52.2	39.3	59.5	47.9	61.6	59	38.7	30.8	5.7
2	28.8	33	26.8	31.9	21.6	26.5	29.2	32	12
3	8.5	13.5	7.4	10.1	6.7	4.6	11.1	14.1	21.7
4	4	5.7	1.9	3.9	2.8	2.8	5.2	6.6	16.9
5	1.4	2	0.8	1.2	1	0.9	1.8	2.7	9.6
6	1.3	1	0.2	0.7	1.1	0.5	0.9	1.4	7
7	0	0.3	0.1	0.2	0.1	0.3	0.7	0.7	4.5
8	0.2	0.1	0	0.1	0.2	0.1	0.1	0.4	3.3
9	0	0.1	0	0	0.2	0.2	0.1	0.1	2
10	0.2	0	0	0	0	0.2	0.1	0.2	1.6
11-15	0	0	0	0	0.1	0.1	0.3	0.3	4.3
16-20	0	0	0	0	0.2	0	0.2	0	1.8
21-50	0	0	0	0	0	0	0	0	3.7
51-100	0	0	0	0	0	0	0	0	0.5
101-150	0	0	0	0	0	0	0	0	0.4
151+	0	0	0	0	0	0	0	0	0

Supplementary Table 3.4.1: Significant *HyPhy* results using the aBSREL model.

Orthogroup	OTU	Omega1	Percent sites	Omega2	Percent sites
OG0000055	M. arenaria A	0	0.9441880363	586.4894174	0.05581196366
	M. incognita A	0	0.9950100058	132.1500633	0.004989994221
OG0000114	M. luci A	0	0.9967664915	5000.0002	0.003233508527
OG0000139	M. javanica B	0	0.9566690114	852.1723014	0.04333098858
	M. luci B	0	0.9358272809	2248.03848	0.06417271908
OG0000142	M. javanica A	0	0.9980010149	1.00E+25	0.001998985113
OG0000243	M. incognita B	0.03807277283	0.7981223056	109.3695148	0.2018776944
OG0000271	M. arenaria A	7.57E-08	0.9872799253	1.00E+25	0.01272007474
OG0000275	M. arenaria A	0	0.9955774441	158.8941006	0.004422555897
	M. incognita A	0.003145523848	0.9913654289	83.76666107	0.008634571133
OG0000286	M. enterolobii	0.03702402973	0.9803211468	225.8029361	0.01967885323
OG0000294	M. arenaria B	0.04133860455	0.9193079971	113.9153009	0.08069200289
	M. incognita A	0	0.9575405461	677.1423468	0.04245945391
	M. javanica B	0.0484536568	0.9778876577	264.8756222	0.0221123423
OG0000297	M. javanica B	0	0.966465248	93.86654736	0.03353475197
OG0000347	M. arenaria B	0	0.9612094327	9.607253735	0.03879056725
	M. incognita B	0.003390029372	0.9966810956	1.00E+25	0.00331890436
OG0000350	M. luci B	0.03560953412	0.9961919831	72.88577165	0.003808016888
	M. arenaria A	0	0.9237395993	1718.319621	0.07626040071
	M. incognita A	0	0.9874583158	170.9352294	0.01254168416
	M. javanica A	0.01633268139	0.917691332	749.8596662	0.08230866803
	M. javanica B	0.04784675648	0.9506907624	53.76395638	0.04930923764
OG0000410	M. luci A	0	0.9900300418	560.3463869	0.009969958222
	M. javanica A	7.35E-08	0.9963528174	1.00E+25	0.003647182624
OG0000419	M. javanica B	4.66E-14	0.9757773053	415.3901417	0.02422269473
	M. arenaria A	7.59E-08	0.9919113311	1.00E+25	0.008088668882
OG0000469	M. javanica B	0	0.9852962441	519.6149362	0.01470375591
OG0000479	M. arenaria A	0	0.9634236629	62.68617287	0.03657633707
OG0000489	M. javanica B	0	0.9871507128	385.0990302	0.01284928722
OG0000503	M. luci A	0	0.6815448977	31.91333202	0.3184551023
OG0000584	M. arenaria B	6.54E-08	0.9552691703	1.00E+25	0.04473082966
OG0000601	M. incognita B	0	0.9790222933	3905.699246	0.02097770672
	M. arenaria B	0	0.9724620337	5711.894055	0.02753796631
OG0000688	M. arenaria A	0	0.967643349	163.0063198	0.03235665096
OG0000724	M. arenaria B	0	0.9346294186	1.00E+25	0.0653705814
	M. incognita A	1.48E-11	0.9707823673	1.00E+25	0.02921763266
OG0000738	M. arenaria A	4.63E-10	0.9540414909	1.00E+25	0.04595850906
	M. arenaria B	6.64E-08	0.990761009	1.00E+25	0.009238990998
OG0000844	M. javanica A	6.66E-08	0.8891240206	1.00E+25	0.1108759794
OG0000900	M. luci B	0	0.9841758082	81.28292051	0.01582419179
	M. arenaria A	0	0.9803041272	1.00E+25	0.01969587275
OG0000911	M. javanica A	0	0.9933030691	1.00E+25	0.006696930915

OG0000946	M. luci B	0	0.9328874286	22262040327	0.06711257143
	M. incognita A	0	0.9935419594	303.3378809	0.006458040601
OG0000949	M. javanica B	0	0.9740102519	8205931110	0.02598974806
	M. arenaria A	0	0.9882676411	20995.40299	0.0117323589
OG0001069	M. javanica A	0	0.9940855179	1.00E+25	0.00591448214
OG0001100	M. incognita A	0	0.9592722769	1718.408712	0.04072772307
OG0001103	M. incognita B	0	0.9878906892	198.6432922	0.01210931075
OG0001126	M. arenaria A	0	0.8364795685	59.45919335	0.1635204315
OG0001181	M. javanica A	0	0.9456323957	1.00E+25	0.05436760432
	M. javanica A	0.03947914695	0.9859729317	477.5693431	0.01402706834
OG0001213	M. incognita B	0	0.9895651801	1.00E+25	0.01043481986
OG0001353	M. arenaria A	0	0.9970640932	2356.813417	0.002935906757
	M. arenaria B	0.02766212598	0.9641841263	1.00E+25	0.03581587371
OG0001356	M. arenaria B	0	0.9552823454	83943292743	0.04471765463
OG0001450	M. javanica A	0	0.9833178034	88.66356021	0.01668219662
OG0001497	M. javanica B	0	0.9803864323	1.00E+25	0.0196135677
	M. incognita A	0	0.9772753059	1.00E+25	0.02272469405
OG0001498	M. javanica A	2.51E-10	0.9604581504	1.00E+25	0.03954184963
	M. arenaria A	7.99E-08	0.9978929081	1.00E+25	0.002107091856
	M. incognita A	0	0.98837807	194.2877339	0.01162193
OG0001509	M. javanica B	0.02018148139	0.9880165446	461.3813854	0.01198345541
OG0001517	M. incognita A	0	0.9401786803	1.00E+25	0.05982131973
OG0001525	M. javanica B	0	0.9882164712	172472411.2	0.01178352883
OG0001531	M. arenaria A	0.01810523088	0.9759246733	215.6754624	0.02407532668
	M. arenaria A	0	0.9829125413	17036.59454	0.01708745871
OG0001542	M. arenaria A	0	0.9808361824	592779.6723	0.01916381763
OG0001718	M. arenaria B	0	0.8420994336	1.00E+25	0.1579005664
OG0001735	M. incognita A	0	0.9896417774	2905.111739	0.0103582226
OG0001764	M. luci B	0.03571956181	0.992629108	5711.751651	0.007370891981
OG0001775	M. arenaria A	0	0.9964910802	692.0733478	0.003508919804
	M. incognita A	0	0.9847383768	1.00E+25	0.01526162321
OG0001783	M. luci A	0.01975595014	0.8618541603	4999.999742	0.1381458397
	M. arenaria A	0	0.9298555098	1151373615	0.07014449024
	M. arenaria B	0	0.9807169729	4192.878222	0.01928302709
	M. incognita A	0	0.9860567498	1.00E+25	0.01394325019
OG0001799	M. arenaria A	0	0.9675905575	1.00E+25	0.0324094425
OG0001810	M. javanica B	0	0.9907966295	19130145091	0.009203370484
	M. javanica A	1.09E-09	0.9167690476	1.00E+25	0.08323095241
OG0001815	M. javanica B	0	0.9330195488	63.25833799	0.06698045121
OG0001845	M. luci B	0.02893810942	0.9943568487	1.00E+25	0.005643151271
OG0001858	M. javanica B	0	0.9783623824	412.2356453	0.02163761761
OG0001859	M. javanica A	0.09214729051	0.7656995747	304.5610254	0.2343004253
	M. arenaria A	0	0.9943163343	1.00E+25	0.005683665734
OG0001869	M. javanica A	0.04403968962	0.991013516	1.00E+25	0.008986483952
OG0001871	M. arenaria B	0	0.9873473365	1.00E+25	0.01265266351

	M. incognita A	0	0.9965075014	479.715146	0.003492498641
	M. javanica B	0	0.9562261716	226.4484829	0.04377382836
OG0001893	M. incognita B	0	0.9905633196	17601.0612	0.009436680444
OG0001893	M. javanica A	0	0.9880435023	1.00E+25	0.01195649775
OG0001906	M. arenaria B	7.09E-08	0.9880376038	1.00E+25	0.01196239618
OG0002268	M. incognita B	0.04583722131	0.9570040261	1.00E+25	0.04299597393
	M. javanica A	0	0.8539192371	1.00E+25	0.1460807629
OG0002282	M. luci B	0	0.9970767319	42677460.57	0.002923268138
OG0002312	M. javanica B	0.02872047465	0.7628242811	3708.305244	0.2371757189
OG0002363	M. arenaria A	0.00750953	0.8931876681	109.7279659	0.1068123319
OG0002363	M. incognita B	0	0.9886116849	1.00E+25	0.01138831509
	M. javanica A	0.04303734598	0.973108822	42.01521613	0.02689117796
OG0002399	M. luci B	0	0.9851072595	1.00E+25	0.01489274046
	M. arenaria B	0	0.9466073998	992.3244398	0.05339260021
	M. incognita B	6.73E-08	0.9929182737	8085457140700	0.007081726288
	M. luci B	0.07154758362	0.9373884055	18.91407987	0.06261159447
OG0002421	M. arenaria A	0	0.9531474797	1.00E+25	0.04685252035
	M. arenaria B	0	0.8206921494	2181.131337	0.1793078506
	M. incognita A	0.0315761962	0.996765398	5000	0.003234601985
OG0002472	M. javanica A	0	0.9752452962	1.00E+25	0.02475470377
OG0002500	M. javanica A	0	0.9377227222	1.00E+25	0.06227727779
OG0002501	M. arenaria A	0	0.9909761533	52.50433855	0.00902384668
OG0002504	M. javanica B	0.03334215146	0.8165486244	202.7292141	0.1834513756
OG0002524	M. arenaria A	0	0.9736171381	1262.722062	0.02638286188
OG0002525	M. arenaria B	0	0.9729442856	1.00E+25	0.02705571444
OG0002530	M. javanica B	0	0.9453459699	181.9387159	0.05465403014
OG0002541	M. javanica A	0.0006865495722	0.8839617455	1.00E+25	0.1160382545
	M. incognita A	0	0.9906898372	5000	0.009310162805
OG0002543	M. luci B	0	0.9554257886	5003.085146	0.0445742114
OG0002595	M. luci A	0	0.9967262897	182.734044	0.003273710335
OG0002599	M. luci B	0	0.9521169397	168.3389394	0.0478830603
	M. arenaria A	1	0.969689435	1.00E+25	0.03031056505
OG0002617	M. luci B	0.02472957778	0.9757885515	1.00E+25	0.02421144845
	M. arenaria B	0.04184566714	0.9622868015	149.9296138	0.03771319846
	M. incognita A	0	0.975793112	1.00E+25	0.02420688798
OG0002619	M. incognita B	0	0.9672287633	1.00E+25	0.03277123671
OG0002996	M. javanica A	0.0005419552596	0.946159756	1.00E+25	0.05384024405
OG0003060	M. javanica A	7.06E-08	0.9814443276	1.00E+25	0.01855567238
OG0003068	M. javanica B	7.35E-08	0.9930996728	10097757012041	0.006900327155
OG0003116	M. luci A	0	0.9974396264	1.00E+25	0.002560373606
OG0003119	M. arenaria B	6.25E-08	0.9859327401	1.00E+25	0.0140672599
OG0003182	M. javanica B	0	0.8353008759	199480189.4	0.1646991241
	M. javanica A	0	0.9823391059	2499	0.01766089407
OG0003209	M. javanica A	0	0.9829504308	27790.584	0.01704956916
OG0003225	M. javanica B	0	0.9644765693	6403.583881	0.0355234307
	M. incognita A	6.51E-08	0.9805453743	14844038462	0.01945462575

	M. javanica A	0.00078196601	0.9251197083	1.00E+25	0.07488029172
OG0003278	M. luci A	0	0.9585256791	129.5090816	0.04147432089
OG0003279	M. incognita A	0.00078196601	0.9777305021	1.00E+25	0.02226949791
OG0003348	M. javanica B	0	0.9698813278	6538.830153	0.0301186722
	M. arenaria B	0.01920519693	0.9640889527	983.0493708	0.03591104733
OG0003350	M. incognita B	0.04902206505	0.9950333436	550.8070978	0.004966656443
	M. javanica A	7.26E-08	0.9745735022	1.00E+25	0.02542649781
OG0003917	M. luci A	0	0.9496645679	68.92141978	0.05033543205
OG0003952	M. incognita B	6.62E-08	0.9767034061	1.00E+25	0.0232965939
	M. javanica B	0	0.9922516152	1.00E+25	0.007748384774
OG0003978	M. javanica A	0	0.9407339744	1.00E+25	0.05926602557
	M. javanica A	0	0.9936062102	9662.16918	0.006393789845
OG0003986	M. javanica B	0	0.9933490092	3715.223704	0.006650990784
	M. arenaria A	0	0.9843609051	1.00E+25	0.01563909485
OG0003986	M. javanica A	0.0396723195	0.9641643668	66.29679003	0.03583563316
OG0004037	M. arenaria A	0	0.9788387789	1.00E+25	0.02116122111
	M. javanica A	0.02472204233	0.9867966177	1.00E+25	0.01320338234
OG0004098	M. javanica A	0	0.9913278116	2000.734934	0.008672188449
OG0004178	M. incognita A	0	0.8754031635	159.176141	0.1245968365
OG0004179	M. javanica B	0	0.9687351258	165837977.5	0.03126487423
	M. incognita A	0	0.9828460425	1568.582457	0.0171539575
	M. incognita B	0.01342584521	0.9909555017	540.8873005	0.009044498291
OG0004186	M. javanica B	0	0.9887090013	60.86719239	0.01129099869
OG0004209	M. incognita A	0	0.9773976927	1.00E+25	0.02260230725
OG0004256	M. luci A	0	0.9528533706	1.00E+25	0.04714662944
	M. javanica A	0	0.9948102771	987.6184728	0.005189722935
OG0005199	M. javanica B	0	0.9973723628	1296.034846	0.002627637228

Supplementary Table 4.1: Genes detected as disrupted or deleted by the deletion in VW5. Putative function was assigned based on best match to known proteins in the NCBI RefSeq database. Genome-wide coverage ratio of VW4 to VW5 is 4.7:1. Transcript names are taken from annotations published by Winter et al. (2024).

Transcript	Disrupted/deleted in VW5	Ontology (best BLAST hit)	Coverage ratio (VW4:VW5)	Notes
PB.2352	-	Nucleoside phosphorylase domain and Uridine phosphorylase	4.92:1	-
PB.2353	-	-	16.85:1	-
PB.2354	-	Calcium binding EGF domain containing protein	13.66:1	-
PB.2355	-	cyclin domain protein	12.33:1	-
PB.2356	-	Probable protein phosphotase	24.81:1	-
PB.2357 (Gene R)	-	RhoGAP domain-containing protein	16.08:1	-
PB.2358.1	-	Protein CBR-NAS-32	9.86:1	-
PB.2359.1	-	-	13.27:1	-
PB.2360.1	-	Formin	9.32:1	-
PB.2361	-	Hypothetical protein GCK72_010392	12.49:1	-
PB.2362	-	Hypothetical protein GCK72_009756	9.93:1	-
PB.2363	-	-	11.79:1	-
PB.2364	-	-	11.05:1	-
PB.2365	-	Arrestin domain-containing protein	15.19:1	-
PB.2366	-	BTB/POZdomain-containing protein	13.6:1	-
PB.2367	Deleted from subgenome A	Hypothetical protein GCK72_008905	138.29:1	-
PB.2368	Deleted from subgenome A	Hypothetical protein GCK72_008905	N/A	Zero coverage in VW5
PB.2369	-	Potassium voltage gated channel subfamily H member 6	11.61:1	-
PB.2370	-	Endonuclease/Exonuclease/Phosphatase	8.72:1	-
PB.2371	-	Nematode cuticle collagen N-terminal domain containing protein	13.98:1	-
PB.2372	-	Hypothetical protein GCK72_024950	14.29:1	-
PB.2373	-	Globin family profile domain-containing protein	11.44:1	-
PB.2374	-	Uncharacterized protein BM_BM11030	10.85:1	-
PB.2375	-	Uncharacterized protein BM_BM6883	11.07:1	-

PB.2376	-	Protein CBR-STR-131	6.92:1	-
PB.2377	-	-	10.43:1	-
PB.2378	-	uncharacterized protein HO173_005441	10.82:1	-
PB.2379	-	-	7.86:1	-
PB.2380	-	-	7.08:1	-
PB.2381	-	-	6.91:1	-
PB.2382	-	Splicing factor	11.83:1	-
PB.2383	-	Adenylate/Guanylate cyclase catalytic domain protein	8.39:1	-
PB.2384	-	Hypothetical protein LOAG_11276	7.68:1	-
PB.2385	-	Lipase 3 isoform X2	8.81:1	-
PB.2386	-	Hypothetical protein LOAG_11276	8.34:1	-
PB.2387	-	Thryotropin-releasing hormone receptor	8.95:1	-
PB.2388	-	KH domain-containing protein	8.28:1	-
PB.2389	-	Pilus assembly protein	26.24:1	-
PB.2390	Deleted from subgenome A	Integrative alpha pat-2 precursor	22.11:1	Zero coverage
PB.2391	Likely deleted in subgenome A	Hypothetical protein LOAG_16918	77.56:1	Half of exons at 1 coverage, rest at 0
PB.2392	Deleted from subgenome A	-	N/A	-
PB.2393	-	Uncharacterised protein BM_BM7512	8.56:1	-
PB.2394	Disrupted	Hypothetical protein LOAG_16918	20.40:1	Several exons at zero coverage
PB.2395	Likely deleted in subgenome A	-	108.39:1	Coverage is <1
PB.2396	Disrupted	Cathepsin B	10.69:1	One of four exons at zero coverage
PB.2397	Deleted in subgenome A	Cathepsin B-like isoform X1	N/A	Zero coverage of transcript and all exons
PB.2398	-	-	10.26:1	-
PB.2399	-	-	6.31:1	-
PB.2400	Disrupted	Aldehyde dehydrogenase domain-containing protein	7.05:1	One exon at zero coverage
PB.2402	-	Cathepsin B-like isoform X1	9.94:1	-

7.3 Supplementary Methods

2.1 DNA extraction and sequencing

High molecular weight DNA isolation

Two ml of lysis buffer containing 100mM NaCl, 10 mM Tris-HCl pH 8.0, 25 mM EDTA, 0.5% (w/v) SDS and 100µg/ml Proteinase K was added to the tube containing ~1,000,000 flash frozen eggs. Samples were mixed with gentle pipetting and homogenised at room temperature overnight. Lysate was then treated with 20µg/ml RNAse at 370C for 30 minutes. The lysate was cleaned with equal volumes of phenol/chloroform using phase lock gels (Quantabio Cat # 2302830). The DNA was precipitated from the cleaned lysate by adding 0.4X volume of 5M ammonium acetate and 3X volume of ice-cold ethanol. The DNA pellet was washed with 70% ethanol twice and resuspended in an elution buffer (10mM Tris, pH 8.0). Purity of gDNA was accessed using NanoDrop ND-1000 spectrophotometer and 260/280 ratio of 1.9 and 260/230 of 2.29 were observed. DNA yield was quantified using Qubit 2.0 Fluorometer (Thermofisher Scientific, MA).

PacBio HiFi

All PacBio HiFi libraries were prepared the following way. HiFi SMRTbell libraries were constructed using the SMRTbell Express Template Prep Kit v2.0 (Pacific Biosciences, Menlo Park, CA; Cat. #100-938-900) according to the manufacturer's instructions. HMW gDNA was sheared to a target DNA size distribution between 15 kb – 20 kb using Diagenode's Megaruptor 3 system (Diagenode, Belgium; Cat. B06010003). The sheared gDNA was concentrated using 0.45X of AMPure PB beads (Pacific Biosciences, Menlo Park, CA; Cat. #100-265-900) for the removal of single-strand overhangs at 37 °C for 15 minutes, followed by further enzymatic steps of DNA damage repair at 37 °C for 30 minutes, end repair and A-tailing at 20 °C for 10 minutes and 65 °C for 30 minutes, ligation of barcoded overhang adapters v3 at 20 °C for 60 minutes and 65 °C for 10 minutes to inactivate the ligase, then nuclease treated at 37 °C for 1 hour. SMRTbell libraries were purified and concentrated with 0.45X Ampure PB beads for size selection using the BluePippin/PippinHT system (Sage Science, Beverly, MA; Cat #BLF7510/HPE7510) to collect fragments greater than 7-9 kb. HiFi SMRTbell libraries were sequenced at UC Davis DNA Technologies Core (Davis, CA) using one SMRT® Cell 8M Tray (Pacific Biosciences, Menlo Park, CA; Cat #101-389-001), Sequel II sequencing chemistry 2.0, and 30-hour movies each on a PacBio Sequel II sequencer.

PacBio Iso-Seq

cDNA synthesis was prepared using NEBNext Single Cell/Low Input cDNA Synthesis & Amplification Module kit (New England Biolabs Inc., Ipswich, MA; Cat #E6421L) according to manufacturer's instructions with a slight modification to the PCR cycles of 15 cycles. Amplified cDNA samples were purified using 0.86X of ProNex beads (Promega, Madison, WI; Cat #NG2003) for equal molar pooling of barcoded cDNA. The pooled cDNA was constructed into the SMRTbell Express Template Prep Kit v2.0 (Pacific Biosciences, Menlo Park, CA; Cat. #100-938-900) according to the manufacturer's instructions. The enzymatic steps included DNA damage repair at 37 °C for 30 minutes, end repair and A-tailing at 20 °C for 30 minutes and 65 °C for 30 minutes, and ligation of adapters v3 at 20 °C for 60 minutes. The SMRTbell library was purified with 1X ProNex beads. The Iso-Seq SMRTbell library was sequenced at UC Davis DNA Technologies Core (Davis, CA) using one SMRT® Cell 8M Tray (Pacific Biosciences, Menlo Park, CA; Cat #101-389-001), Sequel II sequencing chemistry 2.0, and 30-hour movies each on a PacBio Sequel II sequencer.

Oxford Nanopore

A sequencing library was prepared starting with 2 µg of gDNA using the ligation sequencing kit SQK-LSK109 (Oxford Nanopore Technologies, Oxford, UK) following instructions of the manufacturer except for extended incubation times for DNA damage repair, end repair, ligation, and bead elutions. 30 fmol of the final library was loaded on the PromethION flow cell R9.4.1 (Oxford Nanopore Technologies, Oxford, UK) and the data was collected for seventy-two hours. Base calling was performed real-time on the PromethION compute tower using *MinKNOW 20.06.9* and *guppy v4.0*.

2.2 Draft assemblies

Draft assemblies were generated with different assemblers to discern which assembly method or package would perform better with our data. We applied *Canu*, *FALCON*, *Shasta*, *IPA*, and *HiFiasm* with iteratively differing parameters (Chin et al., 2016; Koren et al., 2017; Shafin et al., 2020; Cheng et al., 2021). All assemblers barring *HiFiasm* produced highly fragmented assemblies in comparison, and *HiFiasm* was chosen as the best applicable assembly method. Using *HiFiasm* and iterating through different parameters, we produced upwards of fifty draft assemblies, of which the primary assembly generated during our fifteenth parameter iteration was deemed the best for downstream analysis.

2.3 Assembly appraisal

All the following analyses and assessments were performed as part of the *asmapp* workflow (Winter, 2022), a genome assembly appraisal workflow that automates these processes.

Mitochondrial detection and analysis

The mitochondrial genome was detected using *BLAST* (Altschul et al., 1990), querying a previously published *M. javanica* mitochondrial genome (ACC: NC_026556.1) against a database generated from our assembly. Mitochondrial annotation was performed using the *MITOS* web server (Bernt et al., 2013).

Core Gene Analysis

A core gene presence and completeness analysis was performed by both *CEGMA* (Parra et al., 2007) and *BUSCOv5* (Simão et al., 2015). *CEGMA* was used with its default settings. *BUSCOv5* was used with the eukaryota_odb10 database. We chose to use the eukaryote dataset due to the lack of *Meloidogyne* representation in the nematode *BUSCO* dataset, although analysis with nematoda_odb10 was also performed (Supplementary Table 2.1).

Contamination analysis

Contaminants were detected using *Blobtools* (Laetsch & Blaxter, 2017). *Blobtools* uses GC%, coverage, and taxonomic assignment to identify contaminant contigs within the assembly. Taxonomy was assigned based on Phylum level matches to the complete NCBI *BLAST* nucleotide database.

Coverage

Coverage information was generated using *samtools coverage* (Li et al., 2009). Plots of coverage are generated from the *samtools coverage* output using several custom *R* scripts.

QUAST

Full descriptive statistics of all assemblies were generated using *QUAST* (Gurevich et al., 2013).

2.4 Annotation

Repeat annotation

A repeat model library was generated by *RepeatModeler* (Smit et al., 2015b). This library was then used to inform *RepeatMasker* to identify and mask repeats in the genome (Smit et al., 2015a).

PacBio Iso-Seq

Following sequencing, Iso-Seq reads were provided as full-length non-chimeric sequences which were transformed, mapped to the assembly, and collapsed using the *IsoSeq3* pipeline from Pacific Biosciences with default settings (PacificBiosciences, 2022).

MAKER3

MAKER3 was used to perform prediction-based annotations (Campbell et al., 2014). An initial run was performed using *MAKER3*'s *Exonorate* module, informed by previously published *M. javanica* annotations (Szitenberg et al., 2017) and the Iso-Seq transcriptional library we generated here. Following this, annotations were validated by *fathom*, and a second *MAKER3* run was initiated using the *SNAP* module (Korf, 2004). Annotations were then validated based on annotation edit distance (AED) and extracted as .gff and .fasta.

2.5 Pairs analysis

Orthology

Candidate homoeologous pairs were identified through shared CDS. Iso-Seq transcripts were queried against the assembly using *BLAST* (Altschul et al., 1990) to find orthologs throughout the assembly, with coordinates and counts of matching pairs recorded. Hits were only counted over 97% similarity.

Duplicated core genes

Scaffolds that shared more duplicated *BUSCO* genes than expected by chance were designated as potential homoeologous pairs. *BUSCO* genes are represented in a single copy, per copy. Scaffold pairs that share duplicates of the same *BUSCO* genes are likely homoeologous.

MASH distance

Candidate pairs were also identified using *MASH* hash mapping (Ondov et al., 2016). A sketch was created of the genome and scaffolds were iteratively compared against it. Non-

zero *MASH* similarity scores to non-self scaffolds suggest sequence similarity and homology.

Nucleotide similarity

Scaffolds that share regions of high similarity are potentially homoeologous. We used a combination of *BLAST* and *Nucmer* to identify for each scaffold in the assembly which other non-self scaffold exhibited the highest level of nucleotide similarity (Marçais et al., 2018).

2.6 Phasing subgenomes

Scaffolds identified as present in homoeologous pairs were phased using a method based on that used by Cerca et al. (2021). This method works on the logic that in a hybrid we would expect to find k-mers unique to each subgenomes' lineage, resulting from genomic divergence accumulated since speciation from a common ancestor. Identification of such sequences in a scaffold enables us to assign it to a subgenomic lineage. First, k-mer spectra for each scaffold was generated using *jellyfish* (Marçais & Kingsford, 2011). After importing these k-mer spectra into *R*, tables are created for each homoeologous pair, with each column containing the counts for each given k-mer observed in the pair. These counts are then filtered for abundance; k-mers must appear in a given scaffold at least 75 times, and pairs must share more than 75 orthologs. A second filter is applied to keep only k-mers that are doubly distributed in one counterpart of the pair over the other, to remove k-mers that are not ancestral. To control for varying scaffold lengths, and therefore total k-mer counts, all k-mer counts passing the previous two filters were converted to binary ratios. Scaffolds were then hierarchically clustered according to presence or absence of k-mers and organised into a cladogram, containing two clusters representing scaffolds assigned to subgenome A or subgenome B. The cladogram was plotted using *R*'s base plotting library. Scaffolds assigned to pairs but not phased by the k-mer method were assigned to a subgenome based on the k-mer based assignment of its counterpart.

2.7 Sequence divergence

Within subgenomes

Allelic divergence within subgenomes was estimated in the following way. For each subgenome, a subset of reads exclusive to a given subgenome was extracted from the total mapped HiFi reads. A k-mer distribution was then generated for these reads, and a histogram file generated. These histograms were interpreted with *Genomescope2* (Ranallo-

Benavidez et al., 2020), which predicted the amount of allelic divergence between each subgenomes' alleles.

Between subgenomes

Sequence divergence between subgenomes was calculated for both whole sequence and CDS regions. For whole sequence similarity, subgenome A and subgenome B assigned scaffolds were aligned with *minimap2* (Li, 2018) using the *–ax asm10* preset, and divergence calculated from the number of aligned bases divided by the number of concordant bases. For CDS alignment, the same method was applied using paired CDS extracted during synteny analysis.

2.8 Synteny analysis

We used *MCScan (Python)* (Tang et al., 2008) to detect synteny between phased subgenomes. Subgenome specific CDS annotations generated from Iso-Seq transcript data and gene predictions identified by *MAKER3* were extracted and transformed into bed format. *LAST* aligner was then used to detect orthologs of each feature on the opposing subgenome. Identified syntenic pairs were grouped into blocks based on proximity, producing an anchor file. We then converted this anchor file into a pseudo-bed format containing coordinates of start and stop positions of each syntenic block. These coordinates were used to plot the synteny using *R*. Scripts are available in our Zenodo repository (doi: 10.5281/zenodo.7858245) and also upon request.