Eight novel diagnostic markers differentiate lineages of the highly invasive myrtle rust pathogen *Austropuccinia psidii*

Zhenyan Luo[1], Jinghang Feng[1], Austin Bird[1], Mareike Moeller[1], Rita Tam[1], Luc Shepherd[1,2], Lydia Murphy[1], Lavi Singh[1], Abigail Graetz[1], Lilian Amorim[3], Nelson Sidnei Massola Júnior[3], M. Asaduzzaman Prodhan[2], Louise Shuey[4], Douglas Beattie[5], Alejandro Trujillo Gonzalez[6], Peri A. Tobias[7], Amanda Padovan[1], Rohan Kimber[8], Alistair McTaggart[9], Monica Kehoe[2], Benjamin Schwessinger[1,†,*] , Thaís R. Boufleur[1,3,†,*]

[1] Research School of Biology, The Australian National University, Canberra, ACT 2601, Australia.

[2] DPIRD Diagnostics and Laboratory Services, Department of Primary Industries and Regional Development; 3 Baron-Hay Court, South Perth, WA 6151, Australia.

[3] Department of Plant Pathology and Nematology, Luiz de Queiroz College of Agriculture, University of São Paulo, Piracicaba, SP 13418-900, Brazil.

[4] Queensland Department of Agriculture and Fisheries, Ecosciences Precinct, 41 Boggo Rd, Dutton Park QLD 4102, Australia.

[5] Department of Health, Australian Government.

[6] Centre for Conservation Ecology and Genomics, Institute for Applied Ecology, University of Canberra, Bruce, Australian Capital Territory, Australia.

[7] School of Life and Environmental Sciences, Faculty of Science, The University of Sydney, Camperdown, New South Wales 2006, Australia.

[8] Crop Sciences, South Australian Research and Development Institute (SARDI), Waite Research Precinct, Urrbrae, SA 5064, Australia

[9] Centre for Horticultural Science, Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Ecosciences Precinct, Dutton Park, Queensland, Australia

Z. Luo, J. Feng, and A. Bird contributed equally to this manuscript.

*B. Schwessinger and T. R. Boufleur share the last authorship.

†Corresponding authors: B. Schwessinger and T. R. Boufleur; E-mail: benjamin.schwessinger@anu.edu.au; thaisboufleur@gmail.com

ABSTRACT

Austropuccinia psidii is the causal agent of myrtle rust in over 480 species within the family Myrtaceae. Lineages of *A. psidii* are structured by their hosts in the native range, and some have success in infecting newly encountered hosts. For example, the pandemic biotype has spread beyond South America, and proliferation of other lineages is an additional risk to biodiversity and industries. Efforts to manage *A. psidii* incursions, including lineage differentiation, relies on variable microsatellite markers. Testing these markers is time-consuming, complex, and requires reference material that is not always readily available. We designed a novel diagnostic approach targeting eight selectively chosen loci including the fungal mating-type *HD* (homeodomain) transcription factor locus. The *HD* locus (*bW1/2-HD1* and *bE1/2-HD2)* is highly polymorphic, facilitating clear biological predictions about its inheritance from founding populations. To be considered as potentially derived from the same lineage, all four *HD* alleles must be identical. If all four *HD* alleles are identical six additional markers can further differentiate lineage identity. Our lineage diagnostics relies on PCR amplification of eight loci in different genotypes of *A. psidii* followed by amplicon sequencing using Oxford Nanopore

Technologies (ONT) and comparative analysis. The lineage-specific assay was validated on four isolates with existing genomes, uncharacterized isolates, and directly from infected leaf material. We reconstructed alleles from amplicons and confirmed their sequence identity relative to their reference. Genealogies of alleles confirmed the variations at the loci among lineages/isolates. Our study establishes a robust diagnostic tool for differentiating known lineages of *A. psidii* based on biological predictions and available nucleotide sequences. This tool is suited to detecting the origin of new pathogen incursions.

Keywords: Myrtaceae, Diagnostics, Oxford Nanopore Technologies, mating-type, Homeodomain genes.

Background

Austropuccinia psidii, the causal agent of myrtle rust in over 480 host species within the Myrtaceae family (Carnegie and Giblin, 2020), is among the world's top ten priority fungal species for biosecurity (Hyde et al., 2018). Its broad host range and rapid adaptability to new environments is a threat to biodiversity and industries (Chock, 2020), especially in regions like Australia and New Zealand, where species of the Myrtaceae family are ecologically dominant and culturally important (Hyde et al., 2018). In eucalypts, for example, losses in volume due to rust severity can vary from 23% to 35% (dos Santos et al., 2020).

Initially described in Brazil (Winter, 1884), *A. psidii* remained limited to the Americas for many decades before spreading to all continents, except Europe and Antarctica (Simpson et al., 2006). In Australia, where the pathogen was first detected in 2010 (Carnegie et al., 2010), only the pandemic biotype (group of organisms with identical genetic constitution) has been reported, and isolates belonging to exotic lineages of *A. psidii* are considered a threat to Australian natural environments and commercial native forests (Mackinson et al., 2020, DAFF, 2023). Disease symptoms and spore morphology are highly similar across isolates of *A. psidii* belonging to different lineages, even in cases with strong host associations (Ferrarezi et al., 2022, Boufleur et al., 2023, Morales et al., 2024). The disease is predominantly caused by the clonal stage of the pathogen, urediniospores. It is characterized by the initial appearance of small chlorotic spots developing into bright orange pustules that generate new infective urediniospores.

Early detection and diagnosis are crucial for tracking, and potentially limiting rust fungal incursions (K. Hussain et al., 2020). Microsatellite markers have been used to differentiate lineages of *A. psidii* (Graça et al., 2013, Stewart et al., 2018), but their application can be time-consuming and complex, particularly on a large scale. In addition, they require reference material to calibrate microsatellite profiles, which are not always readily available for *A. psidii*. The current internationally approved assay to diagnose *A. psidii* is a species-specific qPCR (Quantitative Polymerase Chain Reaction) (Baskarathevan et al., 2016, IPPC 2018), however, the choice of locus lacks the variability needed to clearly differentiate among pathogen lineages (Beenken, 2017, Ahmed et al., 2018, Boufleur et al., 2023). Therefore, there is a need to identify novel target regions

that precisely diagnose different lineages of *A. psidii* for faster and precise action in a biosecurity response.

Mating in fungi is controlled through genes expressed at mating-type (*MAT*) loci (Wilson et al., 2015). In rust fungi, these are two unlinked loci, one contains pheromone precursors and receptors (*P/R*) and the other contains homeodomain (*HD*) transcription factors that are closely linked via a short DNA sequence. The *HD* locus encodes *bW-HD1* and *bE-HD2* genes, which are highly multiallelic in rust fungi (Luo et al., 2024) and many other Basidiomycota (Coelho et al., 2017). These transcription factors form heterodimeric complexes between alleles and regulate cellular development during mating and the fungal life cycle (Wilson et al., 2015, Coelho et al., 2017, Cuomo et al., 2017). The analysis of *A. psidii* genomes confirmed physically unlinked, heterozygous *P/R* and *HD* loci, supporting that mate compatibility in this pathogen is governed by two multiallelic *HD* genes (*bW-HD1* and *bE-HD2*) and a biallelic *P/R* gene (Ferrarezi et al., 2022). Hence, the *HD* locus has high discriminatory power for populations due to its high allelic diversity in rust fungi (Holden et al., 2023, Henningsen et al., 2024, Luo et al., 2024). On the other hand, different lineages can share the same *HD* alleles despite being genetically divergent (Holden et al., 2023, Henningsen et al., 2024). This indicates the need for supplemental markers to the *HD* locus to differentiate lineages based on amplicon sequencing.

The aim of this study was to develop a highly sensitive assay for the detection and identification of *A. psidii* lineages. This assay can be used for monitoring existing incursions/outbreaks, and to help prevent and limit further incursions of exotic lineages. Here we introduce novel primers designed to target eight highly discriminatory loci,

including the two mating-type genes (*bW-HD1* and *bH-HD2*), and six additional single copy orthologs (SOGs) that further increase sensitivity and robustness of lineage calling. These primers are designed to be used in combination with long-read sequencing such as those facilitated by Oxford Nanopore Technologies (ONT).

Material and Methods

HD locus identification and primer design

The *HD* locus of five *A. psidii* isolates was identified based on available reference genomes, including Brazilian isolates belonging to two different lineages MF-1 (from *Eucalyptus grandis*) (PRJNA215767, GCA_000469055.2) and LFNJM1 (unpublished data) from *Syzygium jambos* (Boufleur et al., 2023), and APG1 (unpublished data) from *Psidium guajava* 10/16/2024 12:14:00 PM, along with the Au3 isolate that belongs to the pandemic biotype (Au3_v2) (PRJNA810573, GCA_023105745.1, GCA_023105775.1) (Edwards et al., 2022), and the South African isolate Apsidii_AM, that belongs to the South African biotype (PRJNA480390, GCA_003724095.1) (McTaggart et al., 2018). The *HD* locus containing regions were identified with BLASTx (v.2.15.0) (Johnson et al., 2008) in combination with annotated *bW-HD1* and *bE-HD2 A. psidii* genes, as described by Ferrarezi et al., (2022). As expected for the *HD* locus in dikaryotic genome assemblies, two alleles of each *bW-HD1* and *bE-HD2* gene were retrieved. The alleles of each gene (*bW-HD1* and *bE-HD2*) were aligned separately with MAFFT v.7.490 (Katoh and Standley, 2013), alignment gaps were removed manually, and two Bayesian inference

genealogy trees were generated with BEAST2 (Bouckaert et al., 2014), using JC69 + I + G4 (Jukes and Cantor, 1969) as substitution model, with gamma shape and proportion invariant estimated. TreeAnnotator v.2.7.6 (Drummond and Rambaut, 2009) was used to summarise the posterior sample of trees and visualized with Figtree v.1.4.4 (2016).

Primers were designed manually based on a multiple sequence alignment of contigs containing the *HD* alleles of the pandemic Au3_v2 lineage, the South African isolate Apsidii_AM, and the Brazilian isolate MF-1. Two pairs of degenerate primers were designed to amplify part of *bW-HD1* (HDFor2DG + HDMR1) and *bE-HD2* (HDMF2DG + HDRev2DG) individually, and the combination of the most forward and the most reverse primers was used to amplify the partial region of the *HD* locus spanning both *bW-HD1* and *bE-HD2* (HD1_alt_fwd/HD1_alt_rev) (Table 1). The expected amplicon size was ~1600 bp and ~1400 bp for *bW-HD1* and *bE-HD2,* and ~2400 bp for the *HD* locus.

The identification of lineage-specific loci and corresponding primer designing

To distinguish isolates which share identical alleles at the *bW-HD1* and *bE-HD2* loci, we identified single copy orthologs with Orthofinder v.2.5.5 (Emms and Kelly, 2019) using the predicted proteomes of the individual haploid chromosome scale genome assemblies of four isolates (Au3_v2, MF-1, LFNJM1, APG1, n= 2×4 = 8). Orthogroups containing a single copy orthologous gene (SOG) for each haplotype were selected as candidates for additional diagnostic markers. Candidates and their corresponding 200 bp flanking regions were extracted from the genomes and aligned with ClustalO v.0.1.2 (Sievers et al., 2011). Alignments were assessed for their allelic distance and all

candidates with more than six distinct alleles were retained. ClustalO v.0.1.2 was applied to reformat alignments. PrimalScheme (Quick et al., 2017) command line version v.1.4.1 was used to design primers based on multiple sequence alignments. EMBOSS primersearch v.6.6.0.0 (Rice et al., 2000) was used to confirm primers that bind to conserved regions across all alleles while ensuring uniqueness within the haploid genome assemblies to avoid off-target amplifications. Primers which could capture size polymorphisms between alleles were favored. The code used for candidate identification and primer design is available at: https://github.com/ZhenyanLuo/Apsi-diagnostic.

DNA extraction and amplicon amplification

The designed primers were tested with diverse samples, including four positive controls of *A. psidii*, three single-pustule isolates (LFNJM3, LFNJM4 and LFNJRM1), five field samples (CA, CG, LFNEP1, 3.1 and SYD) and three non-target rust species (*bW-HD1* and *bE-HD2* only see Table 2). Genomic DNA (gDNA) was extracted directly from *A. psidii* urediniospores, from infected leaf material or from urediniospores of non-target rust species (Table 2) with the DNeasy Plant mini kit (Qiagen) according to the manufacturer's instructions. The integrity and quality of the DNA was measured with a Nanodrop spectrophotometer (Thermo Fisher Scientific) and checked by agarose gel (0.8%) electrophoresis stained with SYBR Safe (Thermo Fisher Scientific). The DNA concentration was initially determined using the Qubit 4 (Thermo Fisher Scientific) and adjusted to 25 ng/µL for downstream analysis.

In the first round of PCR tests, the aim was to evaluate the amplification of (non-) target sequences by the designed primers followed by sequence analysis using *A. psidii* and three non-target rust isolates (*bW-HD1* and *bE-HD2* only see Table 2 and Figure 1A-B). PCR was performed on a Mastecycler nexus X2 thermal cycler (Eppendorf). The reaction mixture, with a final volume of 25 µL, included 5 µL of 5X reaction buffer (New England Biolabs, NEB), 0.5 µL of dNTPs [10 mM], 1.25 µL of each primer [10 µM], 0.25 µL of Q5 High-Fidelity DNA Polymerase (NEB), 14.75 uL of Nuclease Free Water (NFW) and 2 µL (up to 50 ng) of template DNA. The PCR amplification had an initial denaturation step at 98 °C for 30 s, followed by 35 cycles of denaturation at 98 °C for 30 s, annealing at 58 °C for 30 s and extension at 72 °C for 30 s, with a final extension step at a 72 °C for 2 mins. Specificity tests were run in duplicate, and PCR products were visualised on 2% agarose gel stained with SYBR safe (Thermo Fisher Scientific).

In the second round of PCR tests, the performance of the primers was assessed against single-spore isolates and field samples (urediniospores and infected leaves) collected in Brazil and Australia (Table 2). Identical reaction conditions, as described above, were used to amplify the *bW-HD1* and *bE-HD2* loci individually (Table 2). The amplification of *bW-HD1* and *bE-HD2* loci was performed using the same PCR conditions as for testing. For amplification of 2,400bp amplicons spanning both *bW-HD1* and *bE-HD2*, the reaction mixture, with a final volume of 50 µL, included 10 µL of 5X reaction buffer (New England Biolabs, NEB), 1 µL of dNTPs [10 mM], 5 µL of each primer [10 µM], 0.5 µL of Q5 High-Fidelity DNA Polymerase (NEB), 23.5 µL of Nuclease Free Water (NFW) and 5 µL (up to 50 ng) of template DNA. The PCR amplification was performed using the following conditions: initial denaturation at 98°C for 30 seconds, followed by a

touch-down phase of 10 cycles consisting of 98°C for 10 seconds, 65°C to 61°C for 20 seconds (decreasing by 1°C per cycle), and 72°C for 90 seconds. This was followed by 25 cycles of 98°C for 10 seconds, 61°C for 20 seconds, and 72°C for 90 seconds. The reaction was completed with a final extension at 72°C for 2 minutes. PCR products were visualized on 2% agarose gel stained with SYBR safe (Thermo Fisher Scientific).

For amplifying the additional SOG markers, the reaction mixture, with a final volume of 25 µL, included 5 µL of 5X reaction buffer (New England Biolabs, NEB), 5 µL of Q5 GC enhancer, 0.25 µL of Bovine Serum Albumin (BSA), 0.5 µL of dNTPs [10 mM], 2.5 µL of each primer [10 µM], 0.25 µL of Q5 High-Fidelity DNA Polymerase (NEB), 4 µL of Nuclease Free Water (NFW), and 5 µL of template DNA. The PCR amplification had an initial denaturation step at 98 °C for 30 s, followed by 35 cycles of denaturation at 98 °C for 30 s, annealing at 64 °C for 20 s and extension at 72 °C for 1 minute, with a final extension step at a 72 °C for 5 mins. PCR products were visualized on 2% agarose gel stained with SYBR safe (Thermo Fisher Scientific).

ONT sequencing

For Oxford Nanopore sequencing, libraries were generated following manufacturer instructions for V14 Ligation Sequencing of amplicons (Native Barcoding Kit V14 96 - SQK-NBD114.96) with modifications as follows. An initial bead clean step was performed using 1.2 x of 2 % Sera-Mag beads to purify the PCR product and 200 fmol of clean DNA carried through to End-prep reaction (Hall et al., 2023). The End-prep reaction was incubated at 20 °C and then 65 °C for 15 min to maximize yield. One µL of the end-

prepped DNA was barcoded with 1 µL of Nanopore Native Barcode using 5 µL Blunt/TA ligase master mix (NEB) in total reaction volume of 10 µL for 20 min at 20 °C. The reaction was stopped by adding 1 µL of EDTA to each ligation reaction. The individually barcoded PCR amplicons were pooled, bead cleaned with 0.6 x volume 2% Sera-Mag beads and two washes of 70% ethanol. The pool of barcoded PCR amplicons was eluted in 21 µL of nuclease-free water. Library preparation was completed according to the manufacturers protocol. Twenty fmol of the barcoded library were loaded on a MinION R10.4.1 Flowcell (FLO-MIN114) and sequencing was ran using a MinION Mk1B sequencing device. Basecalling was performed with Guppy v. 6.4.2 Super High Accuracy mode. All long-read amplicon datasets and consensus sequences were deposited to Zenodo (https://doi.org/10.5281/zenodo.10656657).

De novo reconstruction of HD loci and genealogies

A two-step filtering process was implemented on the base-called sequences. In the initial step, sequencing reads were filtered based on their Phred quality scores, with reads having a mean quality score below 15 being removed with Nanofilt v2.6.0 (De Coster et al., 2018). This ensured that the remaining reads had an average per base accuracy of ≥ 97%. The second filtering step involved selecting sequencing reads with lengths around the expected amplicon length: 1,500-1,800 bp for *bW-HD1* and 1,300-1,500 bp for *bE-HD2*.

The VSEARCH orient algorithm (Rognes et al., 2016) was first applied to orient reads according to the reference, then the clustering algorithm was used to quality control

sequencing reads of each sample using a global identity cut-off of 0.85 for clustering. The analysis code is available on Github (https://github.com/TheRainInSpain/Lineage-Specific-Marker.git). And the raw data is available on (https://doi.org/10.5281/zenodo.10656657). Geneious software (v.2023.2.1) was used to visualize the forward read consensus sequences.

The reconstructed amplicon and reference sequences were aligned (Table 2) with MUSCLE (v.5.1) (Edgar, 2004), and genealogical trees were reconstructed using BEAST2 (Bouckaert et al., 2014) with same method described above.

Downstream analysis of secondary loci and genealogies

The base-called sequencing reads were filtered with Nanofilt v.2.6.0 using the same parameter described above. Filtered reads were oriented by applying VSEARCH orient using the pandemic isolate (Au3) target region as reference for each amplicon. The VSEARCH clustering algorithm was used to cluster reads by size with a high identity threshold 0.99. The consensus sequences derived from clusters were aligned to remove duplications. Reads were then mapped back to consensus sequence with BWA-MEM2 for verification. Genealogies of each amplicon were inferred by BEAST2 as described above for *HD* alleles.

To perform Principal component analysis (PCA) for differentiating isolates with four identical *bW-HD1* and *bE-HD2* alleles, we included only four secondary loci (OG4974, OG7530, OG9632, OG9774) which were capable of amplifying at least two alleles per sample. We used the DistanceCalculator from Bio.Phylo to calculate pairwise distances

between amplicons. For samples without a reference genome, amplicons were pseudo-phased based on the haplotype of the reference amplicon with the minimum distance to the query. Amplicons belonging to the same isolate were then merged in the following order: OG4974 hapA, OG4974 hapB, OG7530 hapA, OG7530 hapB, OG9632 hapA, OG9632 hapB, OG9774 hapA, and OG9774 hapB. OneHotEncoder (Pedregosa et al., 2011) was utilized to transform the alignment into a numeric array with five categories (A, T, C, G, -). PCA was then applied to reduce dimensionality and visualize the main sources of variance in the data.

Results

Specificity of the diagnostic assay achieved through HD loci amplification and ONT sequencing

During the initial phase of primer testing, *HD* amplicon PCRs were performed on positive controls of *A. psidii* (Au3, MF-1, APG1 and LFNJM1) and negative controls with non-target rust species (*Miyagia psudosphaeria, Puccinia striiformis* f. sp. *tritici, P. graminis*, *P. triticina, Thekopsora minima*). In addition, we focused on the individual PCR amplicons (*bW-HD1* and *bE-HD2*) because the amplification of the full-length locus was not robust enough across samples and technical replicates. The sizes of the amplicons were as expected in the positive control samples, being of ~1600 bp for *bW-HD1*, ~1400 bp for *bE-HD2*, and ~2400 kb for the full HD locus (Table 2). In addition, we observed bands of variable sizes in some of our technical repeats of non-target species. All samples

were sequenced with our ONT amplicon sequencing workflow because our assay does not rely exclusively on the PCR amplification but requires that the amplified sequences match the *A. psidii HD* sequences in a genealogical framework.

None of the non-target species amplicon sequences matched the full-length *A. psidii HD* sequences. For each *A. psidii* isolate, four *HD* alleles (2 × *bW-HD1* and 2 × *bE-HD2*) were reconstructed based on ONT sequencing results, as expected for dikaryotic organisms. Isolates were considered identical if their four *HD* alleles had >99.9% of pairwise identity. The applicability of the lineage diagnostic test was confirmed by comparing the *de novo* reconstructed consensus sequences derived from ONT amplicon sequencing with the *in silico* derived *bW-HD1/bE-HD2* amplicon sequences based on available reference genomes (APG1, MF-1, LFNJM1, Au3 and Apsidii_AM). All *de novo* reconstructed ONT amplicon sequences clearly grouped with their respective *HD* alleles obtained from reference genomes (Table 2, Figure 1A-B). Moreover, the reference trees for *bW-HD1* and *bE-HD2* revealed that the Brazilian isolates and the South African isolate carry at least two clearly distinct alleles when compared against the pandemic lineage (Figure 1A-B), corroborating variations previously described using microsatellite markers (Graça et al., 2013, Roux et al., 2016, Stewart et al., 2018).

Primers targeting the HD region distinguished different lineages of A. psidii

The designed primers successfully amplified individual *HD* loci of DNA extracted from different sources of field samples, including urediniospores (CA and CG) and infected leaf material (LFNEP1, SYD and 3.1) (Table 2). As observed previously, the full

HD locus amplification was not possible for the APG1 isolate and the field samples (Table 2). The reconstructed *bW-HD1/bE-HD2* amplicons and reference sequences were aligned. Sequences with a pairwise identity >99.9% were considered the same allele across different isolates. Our results revealed that isolate SYD and isolate 3.1, collected from field samples in Australia in 2022 and Brazil in 2023, most likely belongs to the Au3 group, corresponding to the pandemic lineage because all four *HD* alleles clearly grouped with those derived from the pandemic reference isolate (Figure 1C-D). In contrast, other isolates collected from Brazil, belong to different lineages having at least two different *HD* alleles from the pandemic lineage (Figure 1C-D).

Isolates LFNJM1, LFNJM4, LFNNJRM1, LFNEP1 and MF-1 belong to the MF-1/LFNJM1 group, which includes isolates that infect *Eucalyptus* sp., *S. jambos*, *S. samarangense* and *Eugenia stipitata.* Isolates APG1 and LFNJM3 belong to the APG1 group, which infects *P. guajava* and *S. jambos.* Isolates CG and CA did not group with other isolates (Figure 1C-D).

Notably, the 3.1 leaf sample collected from Brazil in 2023 showed identical *bW-HD1* and *bE-HD2* alleles to the pandemic lineage, necessitating further testing with additional secondary diagnostic amplicons.

Secondary diagnostic amplicons allow more sensitive lineage calling

We aimed to further differentiate samples which share identical *HD* alleles e.g. Au3 group, MF1/LFNJM1 group and APG1 group. For this purpose, we identified six additional secondary diagnostic amplicons which target highly variable regions of SOGs (Table 3).

Primer pairs for amplicons OG4974, OG9632, OG9774 amplified two alleles for all samples. Other amplicons recovered two alleles in all samples except for OG7530 amplified three alleles for CG, OG5363 amplified only one allele for LFNJM3, and OG7071 amplified one allele for CG. The genealogical trees show that samples which shared identical *HD* alleles have similar alleles of the additional tested amplicons (Supplementary Figure 2A-F, Supplementary Figure 3-4). However, a deletion specific to the OG9774 allele of Au3 and SYD, and an insertion of a tandem repeat specific to the OG7530 allele of MF1, LFNJM4 and LFENP1 allowed further differentiation of samples which share identical *HD* alleles (Supplementary Figure 4A-B).

To generate the PCA plot, OG4974, OG7530, OG9632 and OG9774 which amplified at least two alleles per sample were selected. The OG7530 amplicon amplified three alleles for CG with OG7530-7 being the best supported by the sequencing reads. In contrast, OG7530-2 and OG7530-1 had fewer reads, but their read counts were similar. Therefore, we used two combinations of OG7530 for the CG isolate: OG7530-2 + OG7530-7 and OG7530-1 + OG7530-7. The PCA plot indicates that SYD belongs to the same lineage as the pandemic lineage Au3. Except for OG7530, the other five secondary diagnostic amplicons of the leaf sample 3.1 were identical to the pandemic lineage. The amplicon for OG7530 distinguished the leaf sample 3.1 from the pandemic biotype in the PCA analysis (Supplementary Figure 3-4). This suggests that the leaf sample 3.1 is derived from an isolates that is likely closely related to or the same biotype as the pandemic lineage. CA, LFNJM3 and CG which have unique combination of *HD* alleles are genetically distinct from all isolates with reference genomes including Au3 (Figure 2A-B). Among the samples with the combinations of *bW5-HD1/bE5-HD2* and *bW6-HD1/bE6-*

HD2 (MF-1/LFNM1 group), LFNJM1 and LFNJRM1 belong to one lineage, while LFNEP1 and MF-1 belong to another. LFNJM4 was derived from a different lineage than the other samples.

Discussion

Traditionally, *A. psidii* lineages have been identified using microsatellite markers (Graça et al., 2013, Kaur et al., 2015, Roux et al., 2016, Stewart et al., 2018). In this study, a diagnostic assay targeting the *bW-HD1*, *bE-HD2*, and six additional SOGs of *A. psidii*, coupled with ONT sequencing was developed. This method successfully amplified diverse copies of all eight loci from *A. psidii* isolates collected in Australia and Brazil and is predicted to work for the South African isolate.

Ten alleles for *bW-HD1* and *bE-HD1* were identified within the isolates used in this study (Table 2, Figure 1). These findings are consistent with the allelic diversity observed in other four rust species, *Puccinia coronata* f. sp. *avenaae*, *P. graminis* f. sp. *tritici, P. triticina* and *P. striiformis* f. sp. *tritici,* where estimates range between 6 to 12 alleles for *bW-HD1* and *bE-HD2* (Holden et al., 2023, Henningsen et al., 2024, Luo et al., 2024). While these allele counts likely underestimate the total circulating *HD* alleles within the global *A. psidii* population, our results support previous findings related to host specialization.

In our study, the combination of alleles *bW1/2-HD1* and *bE1/2-HD2* was found in the isolate Au3 (positive control), SYD (field sample) and only one Brazilian leaf sample (3.1). This Au3 belongs to the pandemic biotype lineage (Graça et al., 2013, Kaur et al.,

2015, Roux et al., 2016, Stewart et al., 2018), which was initially associated with the emergence of myrtle rust across the globe, including Australia, where this genetic group appears to be the only biotype present so far (Sandhu et al., 2016, Stewart et al., 2018). The pandemic biotype of *A. psidii* was previously reported in Colombia, South America (Granados et al., 2017), however, as far as we know, this represents the first report of this biotype or a very closely related lineage in Brazil. Further support for this diagnosis should include the isolation of single pustule cultures from close by infected plant material, whole genome sequencing and analysis. The divergence of alleles observed among isolates from *Eucalyptus* sp./*S. jambos* and *P. guajava* aligns with earlier observations of distinct genetic groups specialized to different hosts in South America (Graça et al., 2013, Stewart et al., 2018). Recently, Morales et al., (2023) further elucidated this specialization for isolates LFNJM1 (from *S. jambos*) and APG1 (from *P. guajava*) to their respective original hosts, these are the same isolates used as positive controls in this study (Table 2). As in other systems like *Puccinia striiformis* f. sp. *tritici* and *Puccinia coronata* f. sp. *avenae* there have been indications that *HD* alleles can be shared between populations (Holden et al., 2023, Henningsen et al., 2024).

The genealogical tree of *de novo* reconstructed amplicons confirmed the biological expectation of lineage specific variation among isolates originating from single spores, field samples from different hosts, and distinct geographic locations. The isolate collected from field samples in Australia had all eight alleles matching to the pandemic lineage isolate Au3, whereas in Brazil, the center of origin of the disease, there was a strong association of *HD* allele status with their original host species, as previously observed (Graça et al., 2013, Stewart et al., 2018, Morales et al., 2023).

Using all eight diagnostic markers we can clearly differentiate non-pandemic samples from the pandemic reference isolate (Au3) and Australian field samples. Hence our new diagnostic protocol serves as a valuable tool in detecting new incursions of the pathogen in regions where a single lineage is present, as for example in Australia where only a single incursion has been reported to date. Further, comparing marker allele identities with known reference sequences could link novel incursions with related populations in source regions. This could help in identifying risks in import pathways of this exotic pathogen and improve risk mitigation strategies. In addition, the assay could potentially detect recombination between populations if purified single pustule isolates were analyzed. In the future, we anticipate that the diagnostic test can be refined to detect urediniospores of *A. psidii* from complex samples derived from air-sampling or mixed infections to enable structured targeted surveillance of this pathogen on the ground.

Acknowledgments

The authors thank Danièle Giblot-Ducray and Kelly Hill for the critical feedback and support, Maria Carolina Quecine for providing the MF-1 isolate of *A. psidii*, and Tuan Duong for sharing WGS data of SA isolate of *A. psidii*. We gratefully acknowledge the computational resources provided by the National Computational Infrastructure (NCI) and ANU Merit Allocation Scheme (ANUMAS).

Funding

A. T. Gonzalez, R. Kimber, and B. Schwessinger were supported by an Australian Government grant from the Department of Agriculture, Fisheries and Forestry entitled “Automated air sampling for remote surveillance and high throughput processing of environmental samples for eDNA analyses”. B. Schwessinger and P. Tobias were supported by an ARC Linkage grant LP190100093. São Paulo Research Foundation (FAPESP) Grant 2019/13191-5. T. R. Boufleur was supported by FAPESP Grants 2022/11900-1 and 2021/01606-6.

Literature Cited

2016. *FigTree* [Online]. Available: http://tree.bio.ed.ac.uk/software/figtree/ [Accessed Nov 17th 2023].

AHMED, M. B., SANTOS, K., SANCHEZ, I. B., PETRE, B., LORRAIN, C., PLOURDE, M. B., DUPLESSIS, S., DESGAGNE-PENIX, I. & GERMAIN, H. 2018. A rust fungal effector binds plant DNA and modulates transcription. *Sci Rep,* 8**,** 14718.

BASKARATHEVAN, J., TAYLOR, R., HO, W., MCDOUGAL, R., SHIVAS, R. G. & ALEXANDER, B. 2016. Real-time PCR assays for the detection of *Puccinia psidii*. *Plant Disease,* 100**,** 617-624.

BEENKEN, L. 2017. Austropuccinia: a new genus name for the myrtle rust *Puccinia psidii* placed within the redefined family *Sphaerophragmiaceae* (Pucciniales). *Phytotaxa,* 297**,** 53-61.

BOUCKAERT, R., HELED, J., KÜHNERT, D., VAUGHAN, T., WU, C.-H., XIE, D., SUCHARD, M. A., RAMBAUT, A. & DRUMMOND, A. J. 2014. BEAST 2: a

software platform for Bayesian evolutionary analysis. *PLoS computational biology,* 10**,** e1003537.

BOUFLEUR, T. R., MORALES, J. V., MARTINS, T. V., GONÇALVES, M. P., JÚNIOR, N. S. M. & AMORIM, L. 2023. A diagnostic guide for myrtle rust. *Plant Health Progress,* 24**,** 242-251.

CARNEGIE, A. & GIBLIN, F. 2020. Austropuccinia psidii (myrtle rust). Invasive species compendium. ICABI. https://www. cabi. org/isc/datas heet/45846.

CARNEGIE, A., LIDBETTER, J., WALKER, J., HORWOOD, M., TESORIERO, L., GLEN, M. & PRIEST, M. 2010. Uredo rangelii, a taxon in the guava rust complex, newly recorded on Myrtaceae in Australia. *Australasian Plant Pathology,* 39**,** 463-466.

CHOCK, M. 2020. The global threat of Myrtle rust (*Austropuccinia psidii*): Future prospects for control and breeding resistance in susceptible hosts. *Crop protection,* 136**,** 105176.

COELHO, M. A., BAKKEREN, G., SUN, S., HOOD, M. E. & GIRAUD, T. 2017. Fungal sex: the Basidiomycota. *The fungal kingdom***,** 147-175.

CUOMO, C. A., BAKKEREN, G., KHALIL, H. B., PANWAR, V., JOLY, D., LINNING, R., SAKTHIKUMAR, S., SONG, X., ADICONIS, X. & FAN, L. 2017. Comparative analysis highlights variable genome content of wheat rusts and divergence of the mating loci. *G3: Genes, Genomes, Genetics,* 7**,** 361-376.

DAFF. 2023. *Myrtle rust* [Online]. Available: https://www.agriculture.gov.au/biosecurity-trade/pests-diseases-weeds/plant/myrtle-rust [Accessed 2024 08.21].

DE COSTER, W., D'HERT, S., SCHULTZ, D. T., CRUTS, M. & VAN BROECKHOVEN, C. 2018. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics,* 34**,** 2666-2669.

DOS SANTOS, A. P., GOMES, R. L., FURTADO, E. L. & DE SOUZA PASSOS, J. R. 2020. Quantifying losses in productivity by the rust in eucalypt plantations in Brazil. *Forest Ecology and Management,* 468**,** 118170.

DRUMMOND, A. J. & RAMBAUT, A. 2009. Bayesian evolutionary analysis by sampling trees. *The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing***,** 564-574.

EDGAR, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research,* 32**,** 1792-1797.

EDWARDS, R. J., DONG, C., PARK, R. F. & TOBIAS, P. A. 2022. A phased chromosome-level genome and full mitochondrial sequence for the dikaryotic myrtle rust pathogen, Austropuccinia psidii. *bioRxiv*.

EMMS, D. M. & KELLY, S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome biology,* 20**,** 1-14.

FERRAREZI, J., MCTAGGART, A., TOBIAS, P., HAYASHIBARA, C., DEGNAN, R., SHUEY, L., FRANCESCHINI, L., LOPES, M. & QUECINE, M. 2022. *Austropuccinia psidii* uses tetrapolar mating and produces meiotic spores in older infections on *Eucalyptus grandis*. *Fungal Genetics and Biology***,** 103692.

Geneious 2024.0.4 (https://www.geneious.com).

GLEN, M., ALFENAS, A., ZAUZA, E., WINGFIELD, M. & MOHAMMED, C. 2007. *Puccinia psidii*: a threat to the Australian environment and economy—a review. *Australasian Plant Pathology,* 36**,** 1-16.

GRAÇA, R. N., ROSS-DAVIS, A. L., KLOPFENSTEIN, N. B., KIM, M. S., PEEVER, T. L., CANNON, P. G., AUN, C. P., MIZUBUTI, E. S. & ALFENAS, A. C. 2013. Rust disease of eucalypts, caused by *Puccinia psidii*, did not originate via host jump from guava in B razil. *Molecular ecology,* 22**,** 6033-6047.

GRANADOS, G. M., MCTAGGART, A. R., BARNES, I., RODAS, C., ROUX, J. & WINGFIELD, M. J. 2017. The pandemic biotype of *Austropuccinia psid*ii discovered in South America. *Australasian Plant Pathology,* 46**,** 267-275.

HALL, R. N., JONES, A., CREAN, E., MARRIOTT, V., PINGAULT, N., MARMOR, A., SLOAN-GARDNER, T., KENNEDY, K., COLEMAN, K. & JOHNSTON, V. 2023. Public health interventions successfully mitigated multiple incursions of SARS-CoV-2 Delta variant in the Australian Capital Territory. *Epidemiology & Infection,* 151**,** e30.

HENNINGSEN, E. C., LEWIS, D., NAZARENO, E., HUANG, Y.-F., STEFFENSON, B. J., BOESEN, B., KIANIAN, S. F., STONE, E., DODDS, P. N. & SPERSCHNEIDER, J. 2024. A high-resolution haplotype pangenome uncovers somatic hybridization, recombination and intercontinental migration in oat crown rust. *bioRxiv***,** 2024.03. 27.583983.

HOLDEN, S., BAKKEREN, G., HUBENSKY, J., BAMRAH, R., ABBASI, M., QUTOB, D., DE GRAAF, M.-L., KIM, S. H., KUTCHER, H. R., MCCALLUM, B. D., RANDHAWA, H. S., IQBAL, M., ULOTH, K., BURLAKOTI, R. R. & BRAR, G. S.

2023. Uncovering the history of recombination and population structure in western Canadian stripe rust populations through mating type alleles. *BMC Biology,* 21**,** 233.

HYDE, K. D., AL-HATMI, A. M., ANDERSEN, B., BOEKHOUT, T., BUZINA, W., DAWSON, T. L., EASTWOOD, D. C., JONES, E. G., DE HOOG, S. & KANG, Y. 2018. The world's ten most feared fungi. *Fungal Diversity,* 93**,** 161-194.

IPPC. 2018. DP 26: *Austropuccinia psidii*.

JOHNSON, M., ZARETSKAYA, I., RAYTSELIS, Y., MEREZHUK, Y., MCGINNIS, S. & MADDEN, T. L. 2008. NCBI BLAST: a better web interface. *Nucleic acids research,* 36**,** W5-W9.

JUKES, T. H. & CANTOR, C. R. 1969. Evolution of protein molecules. *Mammalian protein metabolism,* 3**,** 21-132.

K. HUSSAIN, K., MALAVIA, D., M. JOHNSON, E., LITTLECHILD, J., WINLOVE, C. P., VOLLMER, F. & GOW, N. A. 2020. Biosensors and diagnostics for fungal detection. *Journal of Fungi,* 6**,** 349.

KATOH, K. & STANDLEY, D. M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution,* 30**,** 772-780.

KAUR, S., PANESAR, P. S., BERA, M. B. & KAUR, V. 2015. Simple sequence repeat markers in genetic divergence and marker-assisted selection of rice cultivars: a review. *Critical Reviews in Food Science and Nutrition,* 55**,** 41-49.

LUO, Z., MCTAGGART, A. & SCHWESSINGER, B. 2024. Genome biology and evolution of mating-type loci in four cereal rust fungi. *PLoS Genetics,* 20**,** e1011207.

MACKINSON, B., PEGG, G. S. & CARNEGIE, A. J. 2020. Myrtle Rust in Australia–a National Action Plan. Australian Plant Biosecurity Science Foundation.

MCTAGGART, A. R., DUONG, T. A., LE, V. Q., SHUEY, L. S., SMIDT, W., NAIDOO, S., WINGFIELD, M. J. & WINGFIELD, B. D. 2018. Chromium sequencing: the doors open for genomics of obligate plant pathogens. *Biotechniques,* 65**,** 253-257.

MORALES, J. V., BOUFLEUR, T. R., GONÇALVES, M. P., PARISI, M. C., LOEHRER, M., SCHAFFRATH, U. & AMORIM, L. 2023. Differential aggressiveness of *Austropuccinia psidii* isolates from guava and rose apple upon cross-inoculation. *Plant Pathology*.

MORALES, J. V., BOUFLEUR, T. R., GONÇALVES, M. P., PARISI, M. C., LOEHRER, M., SCHAFFRATH, U. & AMORIM, L. 2024. Differential aggressiveness of *Austropuccinia psidii* isolates from guava and rose apple upon cross-inoculation. *Plant Pathology,* 73**,** 915-923.

PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R. & DUBOURG, V. 2011. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research,* 12**,** 2825-2830.

QUICK, J., GRUBAUGH, N. D., PULLAN, S. T., CLARO, I. M., SMITH, A. D., GANGAVARAPU, K., OLIVEIRA, G., ROBLES-SIKISAKA, R., ROGERS, T. F. & BEUTLER, N. A. 2017. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nature protocols,* 12**,** 1261-1276.

RICE, P., LONGDEN, I. & BLEASBY, A. 2000. EMBOSS: the European molecular biology open software suite. Elsevier current trends.

ROGNES, T., FLOURI, T., NICHOLS, B., QUINCE, C. & MAHÉ, F. 2016. VSEARCH: a versatile open source tool for metagenomics. *PeerJ,* 4**,** e2584.

ROUX, J., GRANADOS, G. M., SHUEY, L., BARNES, I., WINGFIELD, M. J. & MCTAGGART, A. R. 2016. A unique genotype of the rust pathogen, *Puccinia psidii*, on Myrtaceae in South Africa. *Australasian Plant Pathology,* 45**,** 645-652.

SANDHU, K., KARAOGLU, H., ZHANG, P. & PARK, R. 2016. Simple sequence repeat markers support the presence of a single genotype of *Puccinia psidii* in Australia. *Plant Pathology,* 65**,** 1084-1094.

SIEVERS, F., WILM, A., DINEEN, D., GIBSON, T. J., KARPLUS, K., LI, W., LOPEZ, R., MCWILLIAM, H., REMMERT, M. & SÖDING, J. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology,* 7**,** 539.

SIMPSON, J., THOMAS, K. & GRGURINOVIC, C. 2006. Uredinales species pathogenic on species of Myrtaceae. *Australasian Plant Pathology,* 35**,** 549-562.

STEWART, J., ROSS-DAVIS, A., GRAÇA, R., ALFENAS, A., PEEVER, T., HANNA, J., UCHIDA, J., HAUFF, R., KADOOKA, C. & KIM, M. S. 2018. Genetic diversity of the myrtle rust pathogen (*Austropuccinia psidii*) in the Americas and Hawaii: global implications for invasive threat assessments. *Forest Pathology,* 48**,** e12378.

WILSON, A. M., GODLONTON, T., VAN DER NEST, M. A., WILKEN, P. M., WINGFIELD, M. J. & WINGFIELD, B. D. 2015. Unisexual reproduction in *Huntiella moniliformis*. *Fungal Genetics and Biology,* 80**,** 1-9.

WINTER, G. 1884. Rabenhorstii Fungi europaei et extraeuropaei exsiccati cura Dr. *G. Winter, Centuria XXXI et XXXII. Hedwigia,* 23**,** 164-172.

The authors declare no conflict of interest.

This manuscript is preprinted at BioRXIV under doi.org/10.1101/2024.02.19.580897

Tables

Table 1. List of primers used in the present study.

Primers	Sequence (5' - 3')	Changes	Amplicon size (bp)
HDFor2DG (forward)	ATACAGTTYAGGTTWTRGCG	C/T, A/T, A/G	~ 1600
HDMR1 (reverse)	GAAAGGAAATATTGCCACT	-	
HDMF2DG (forward)	YGACCGCCTTCCTTTGAG	C/T	~ 1400
HDRev2DG (reverse)	GTGTCSAAGCWACCAAAATC	C/G, A/T	
HD1_alt_fwd (forward)	AGTTGATGARGRAGTGGAAG	A/G	~ 2400
HD1_alt_rev (reverse)	CCARCGATTGACTTGATCACG	A/G	
OG4974 (forward)	CAACAGCAACCGTCTCAACAG	-	~ 1400
OG4974 (reverse)	ATCGGTTGAGTTAGCGCCAC	-	
OG5363 (forward)	AGCATCTAACCAAACCCCATCAC	-	~1500
OG5363 (reverse)	ATGAGGCCGCCTAATAAGCAAG	-	
OG7071 (forward)	CGACATTCACTCCTTTGCTGGT	-	~1700
OG7071 (reverse)	ACAATGTGAGCAAGGAACTGCT	-	
OG7530 (forward)	ACAAGAATCACTGGTGTGGAAGG	-	~1550
OG7530 (reverse)	CTTTGAGTTACAGTTGTGCACCTG	-	
OG9632 (forward)	CGACAGGTATGGGCTAGGAC	-	~1550
OG9632 (reverse)	AAAACTTGCTTCGACGTCCG	-	
OG9774 (forward)	TGTGACGTTTCCACCTCAGG	-	~1600
OG9774 (reverse)	TGCCCGATTTCTCACTCCAC	-	

Table 2. Description of isolates and populations and of *A. psidii* and non-target rusts used in this study

Organism	Code	Source	Country	Amplification[a]			Alleles			
				Full locus	*bW-HD1*	*bE-HD2*	*bW-HD1*		*bE-HD2*	
Genomes										
Austropuccinia psidii	Au3	*Agonis flexuosa*	AUS	NA	NA	NA	*bW1*	*bW2*	*bE1*	*bE2*
A. psidii	MF-1	*Eucalyptus grandis*	BR	NA	NA	NA	*bW5*	*bW6*	*bE5*	*bE6*
A. psidii	APG1 (GM1)	*Psidium guajava*	BR	NA	NA	NA	*bW1**	*bW7*	*bE1*	*bE7*
A. psidii	LFNJM1	*Syzygium jambos*	BR	NA	NA	NA	*bW5*	*bW6*	*bE5*	*bE6*
A. psidii	Apsidii AM	*S. jambos*	SA	NA	NA	NA	*bW3*	*bW4*	*bE3*	*bE4*
Isolates (single-pustule)										
Austropuccinia psidii	Au3[b]	*A. flexuosa*	AUS	+	+	+	*bW1*	*bW2*	*bE1*	*bE2*
A. psidii	MF-1[b]	*E. grandis*	BR	+	+	+	*bW5*	*bW6*	*bE5*	*bE6*
A. psidii	APG1 (GM1)[b]	*Psidium guajava*	BR	-	-	+	NA	NA	*bE7*	NA
A. psidii	LFNJM1[b]	*Syzygium jambos*	BR	+	+	+	*bW5*	*bW6*	*bE5*	*bE6*
A. psidii	LFNJM3	*S. jambos*	BR	+	+	-	*bW1*	*bW7*	*bE1*	*bE7*
A. psidii	LFNJM4	*S. jambos*	BR	+	+	+	*bW5*	*bW6*	*bE5*	*bE6*
A. psidii	LFNJRM1	*S. samarangense*	BR	+	+	+	*bW5*	*bW6*	*bE5*	*bE6*
Populations (urediniospores/ field samples)										
A. psidii	CA	*Plinia edulis*	BR	+	+	+	*bW8*	*bW9*	*bE8*	*bE9*

A. psidii	CG	*Eugenia dysenterica*	BR	+	+	+	*bW3*	*bW10*	*bE3*	*bW10*
Populations (infected leaves/ field samples)										
A. psidii	LFNEP1	*Eugenia stipitata*	BR	+	+	+	*bW5*	*bW6*	*bE5*	*bE6*
A. psidii	3.1	*Pimenta dioica*	BR	+	+	+	*bW1*	*bW2*	*bE1*	*bE2*
A. psidii	SYD	*Melaleuca quinquenervia*	AUS	-	+	+	*bW1*	*bW2*	*bE1*	*bE2*
Non-target rust species										
Miyagia psudosphaeria	-	-	-	NS	NP	NA	NA	NA	NA	NA
Puccinia triticina	-	-	-	NS	NS	NA	NA	NA	NA	NA
P. graminis f. sp. *avenae*	-	-	-	NS	NS	NA	NA	NA	NA	NA
P. striiformis f. sp. *tritici*	-	-	-	NS	NS	NA	NA	NA	NA	NA
Thekopsora minima	-	-	-	-	NS	NA	NA	NA	NA	NA

[a] Conventional polymerase chain reaction (PCR) with the three sets of primers developed in the present study. Full *HD* locus, HD1_alt_fwd (forward)/ HD1_alt rev (reverse) amplifies a fragment of 2400 kb; *bW-HD1,* HDFor2DG/HDMR1 amplifies a fragment of 1500 kb; *bE-HD2,* HDMF2/HDRev2DG amplifies a fragment of 1400 kb; +, positive; -, negative.

[b] *A. psidii* and non-target rust isolates used to confirm the specificity of the designed primers.

“NS” is defined as non-specific amplification in at least one of the technical replicates, “NA” as not available, "-" as no amplification, “+” as amplification of the expected size amplicon product.

Table 3. Classification of secondary diagnostic markers of *Austropuccinia psidii*

Isolate code	OG4974		OG5363		OG7071		OG7530		OG9632		OG9774	
	hapA	***hapB***	***hapA***	***hapB***	***hapA***	***hapB***	***hapA***	***hapB***	***hapA***	***hapB***	***hapA***	***hapB***
Genomes												
Au3	OG4974-1	OG4974-2	OG5363-1	OG5363-2	OG7071-1	OG7071-2	OG7530-1	OG7530-2	OG9632-1	OG9632-2	OG9774-1	OG9774-2
MF-1	OG4974-4	OG4974-5	OG5363-4	OG5363-5	OG7071-4	OG7071-5	OG7530-5	OG7530-6	OG9632-3	OG9632-4	OG9774-2	OG9774-3
APG1 (GM1)	OG4974-3	OG4974-6	OG5363-3	OG5363-6	OG7071-3	OG7071-1	OG7530-3	OG7530-4	OG9632-5	OG9632-7	OG9774-4	OG9774-5
LFNJM1	OG4974-4	OG4974-5	OG5363-4	OG5363-5	OG7071-4	OG7071-5	OG7530-5*	OG7530-6	OG9632-3	OG9632-4	OG9774-2	OG9774-3
Isolates (single-pustule)												
Au3[b]	OG4974-1	OG4974-2	OG5363-1	OG5363-2	OG7071-1	OG7071-2	OG7530-1	OG7530-2	OG9632-1	OG9632-2	OG9774-1	OG9774-2
MF-1[b]	OG4974-4	OG4974-5	OG5363-4	OG5363-5	OG7071-4	OG7071-5	OG7530-5	OG7530-6	OG9632-3	OG9632-4	OG9774-2	OG9774-3
APG1 (GM1)[b]	OG4974-3	OG4974-6	OG5363-3	OG5363-6	OG7071-3	OG7071-1	OG7530-3	OG7530-4	OG9632-5	OG9632-7	OG9774-4	OG9774-5
LFNJM1[b]	OG4974-4	OG4974-5	OG5363-4	OG5363-5	OG7071-4	OG7071-5	OG7530-5*	OG7530-6	OG9632-3	OG9632-4	OG9774-2	OG9774-3
LFNJM3	OG4974-7	OG4974-6*	OG5363-4*	NA	OG7071-6	OG7071-6	OG7530-2	OG7530-4	OG9632-1*	OG9632-8	OG9774-4	OG9774-5
LFNJM4	OG4974-4	OG4974-5*	OG5363-4	OG5363-5	OG7071-4	OG7071-5	OG7530-5	OG7530-6	OG9632-3	OG9632-4	OG9774-2	OG9774-3
LFNJRM1	OG4974-4	OG4974-5	OG5363-4	OG5363-5	OG7071-4	OG7071-5	OG7530-5*	OG7530-6	OG9632-3	OG9632-4	OG9774-2	OG9774-3
Populations (urediniospores/ field samples)												
CA	OG4974-1*	OG4974-8	OG5363-7	OG5363-8	OG7071-7	OG7071-2	OG7530-1/OG7530-2	OG7530-7	OG9632-9	OG9632-6	OG9774-1	OG9774-3
CG	OG4974-5*	OG4974-5	OG5363-5*	OG5363-5	NA	OG7071-5	OG7530-2	OG7530-2	OG9632-8	OG9632-4	OG9774-3	OG9774-3
Populations (infected leaves/ field samples)												
LFNEP1	OG4974-4	OG4974-5	OG5363-4	OG5363-5	OG7071-4	OG7071-5	OG7530-5	OG7530-6	OG9632-3	OG9632-4	OG9774-2	OG9774-3
3.1	OG4974-1	OG4974-2	OG5363-1	OG5363-2	OG7071-1	OG7071-2	OG7530-1	OG7530-2	OG9632-1	OG9632-2	OG9774-1*	OG9774-2
SYD	OG4974-1	OG4974-2	OG5363-1	OG5363-2	OG7071-1	OG7071-2	OG7530-1	OG7530-2	OG9632-1	OG9632-2	OG9774-1*	OG9774-2

* The '*' symbol is added to an allele if it differs by at least 3 base pairs from other alleles.

Figure legends

Figure 1: Bayesian inference genealogical trees reconstructed from *HD* (homeodomain) genes of *Austropuccinia psidii*. Genealogical trees reconstructed from the alignment of two *bE-HD2* **(A, C)** or two *bW-HD1* **(B, D)** alleles per dikaryotic genome assembly of reference isolates **(A-B),** and with consensus sequences for each allele *de novo* reconstructed from ONT long-read amplicons of partial *bE-HD2* (C) and *bW-HD1* (D). Lineage identification is indicated by the combination of different colors in the outer ring of each tree. Reference sequences are indicated with (ref). Bayesian posterior probabilities are shown on the nodes, posterior value of leaves which shared identical sequences are removed. The scale bar represents 0.05 substitutions per site.

Figure 2: Principal Component Analysis (PCA) of four amplicons (OG4974, OG7530, OG9632 and OG9774) show distinguishable differences between samples. Each point represents a single sample and the spatial relationships between points indicate similarities or differences within alignments. (**A**) PCA plot of alignment with CA has combination of OG7530-1 and OG7530-7. (**B**) PCA plot of alignment with CA has combination of OG7530-2 and OG7530-7. 3.1 which has identical *HD* alleles as the pandemic lineage segregate from the pandemic lineage sample, reference and SYD isolate. JM1 and JRM1 clustered as one point, while LFNEP1 and MF1 clustered into another. JM4 was derived from a different lineage than the other samples.

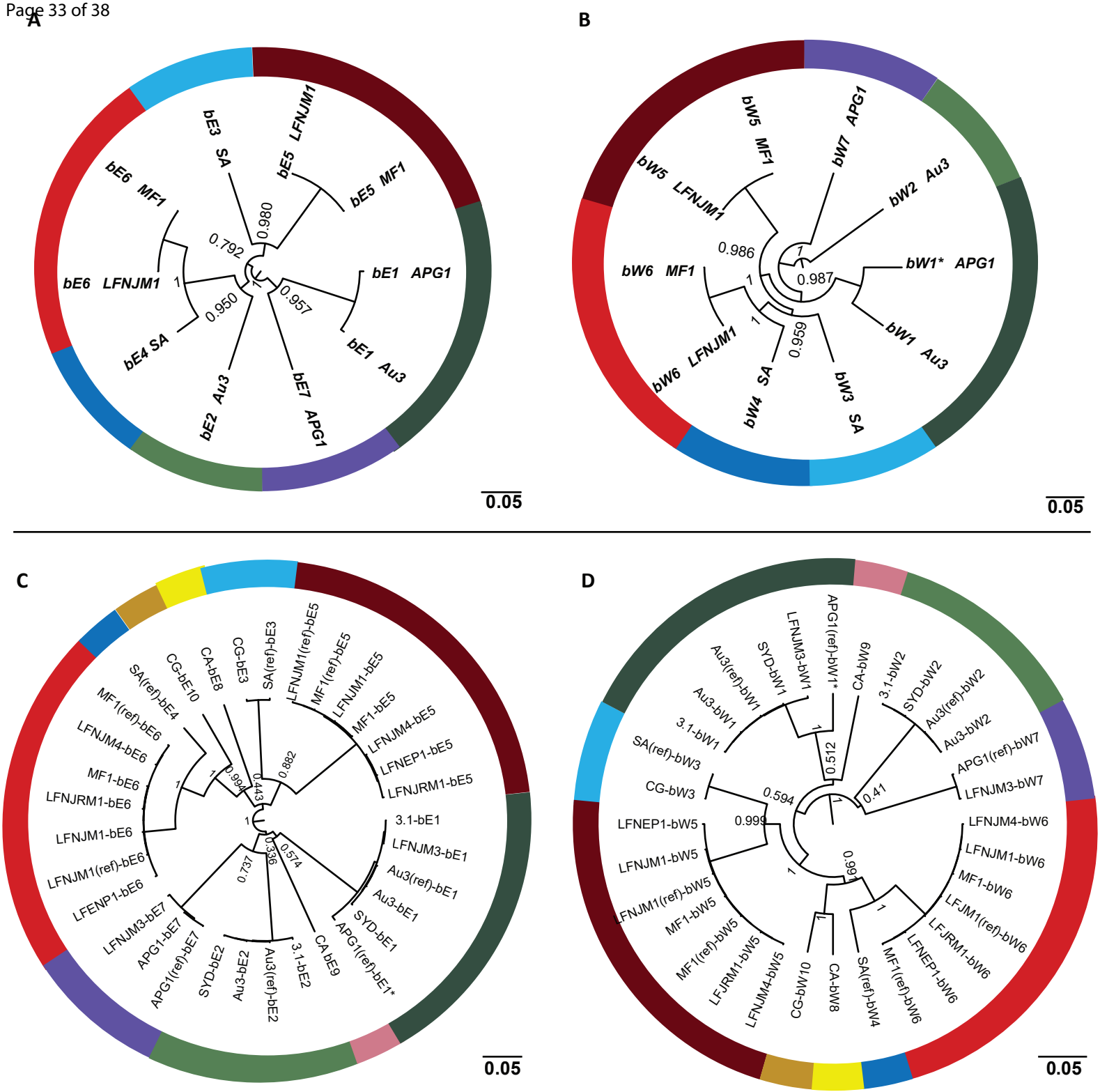
Supplementary Figure 1: Multiple sequence alignment of *bE-HD2* (A) and *bW-HD1* (B) alleles derived from reference genome assemblies and *de novo* reconstructed partial *bE-HD2* (C) and *bW-HD1* (D). *bE-HD2* (A) and *bW-HD1* (B) CDS were aligned using MAFFT and visualized in Geneious Prime 2024.0.4. The *bE-HD2* (C) and *bW-HD1* (D) CDS regions were extracted from amplicons. Each large rectangle represents an individual nucleotide sequence. Grey regions

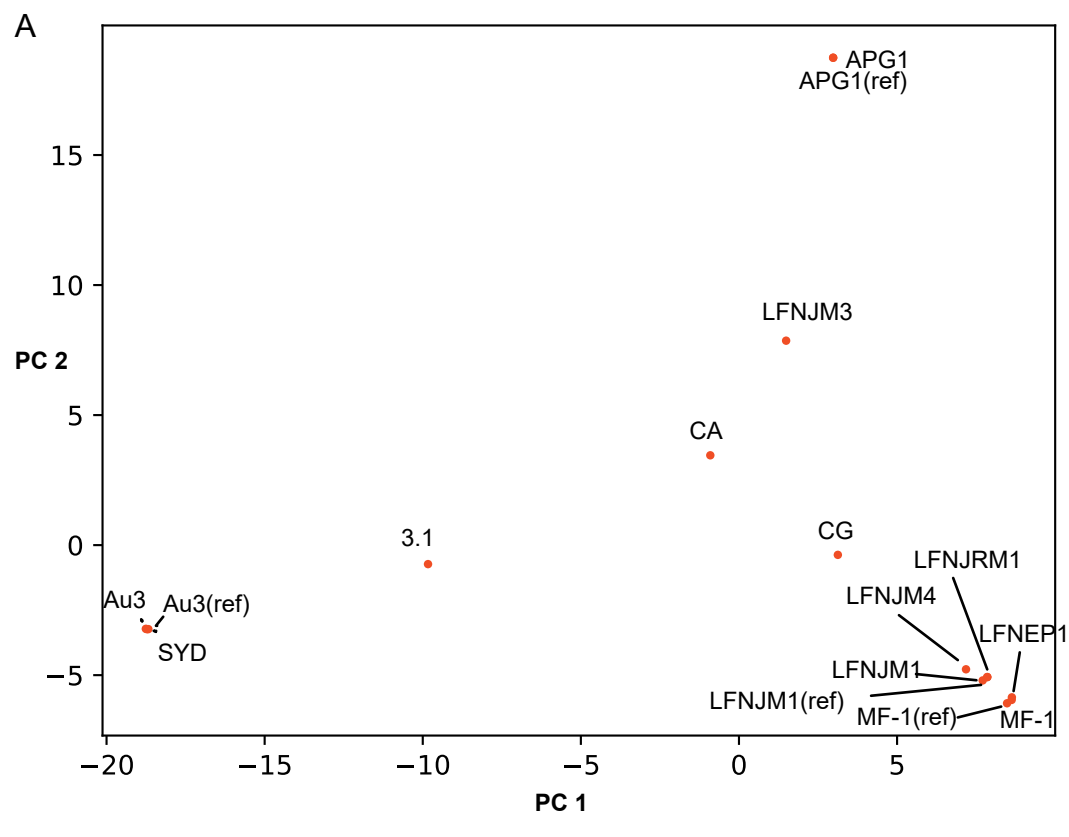
represent nucleotides identical to the consensus sequence, whereas black or coloured lines represent nucleotide difference between alleles.

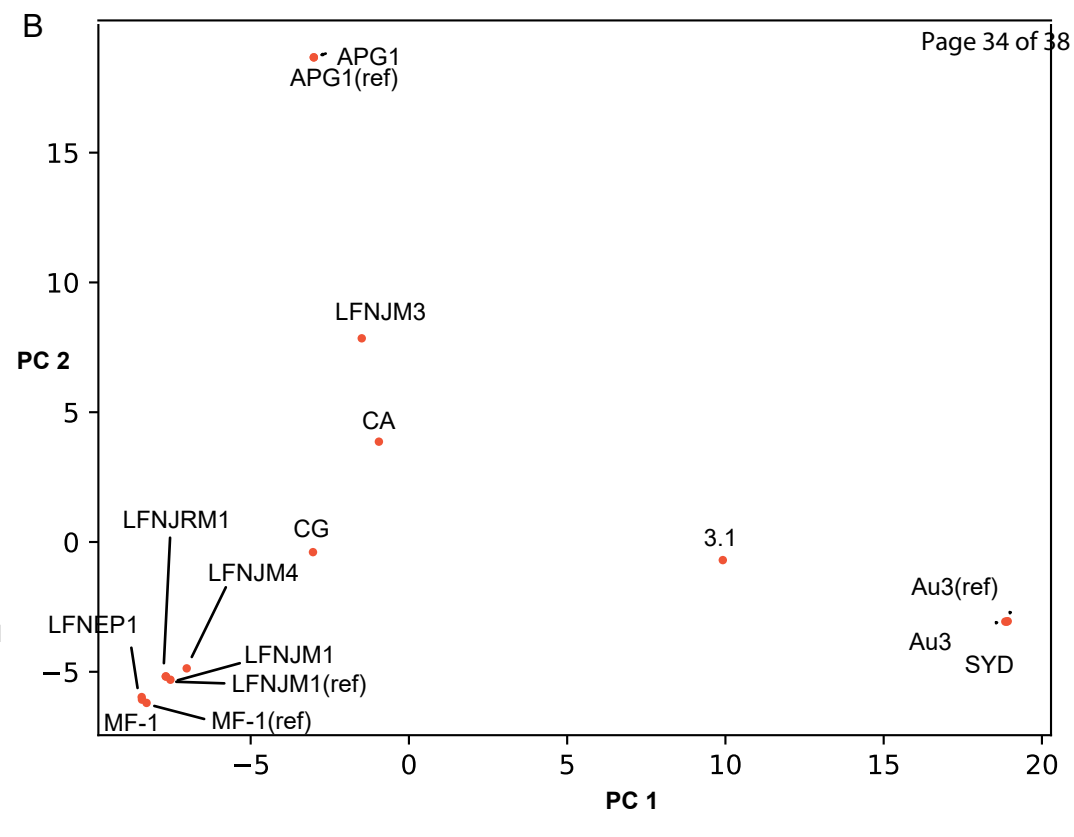
Supplementary Figure 2: Bayesian inference genealogical trees reconstructed from six secondary diagnostic amplicons OG4974 (A), OG5363 (B), OG7071 (C), OG7530 (D), OG9632 (E) and OG9774 (F). Bayesian posterior probabilities are shown on the nodes, posterior value of leaves which shared identical sequences are removed. The scale bar represents 0.05 substitutions per site. Reference sequences are indicated in bold font style. The '*' symbol is added to an allele if it differs by at least 3 base pairs from other alleles.

Supplementary Figure 3: Multiple sequence alignment of *de novo* reconstructed OG4974 (A), OG5363 (B), OG7071 (C) and OG9632 (D). Consensus sequences were aligned using MAFFT and visualized in Geneious Prime 2024.0.4. Each large rectangle represents an individual nucleotide sequence. Grey regions represent nucleotides identical to the consensus sequence, whereas black narrow lines represent nucleotide diversity across alleles, narrow lines represent gaps, black blocks represent inspections.

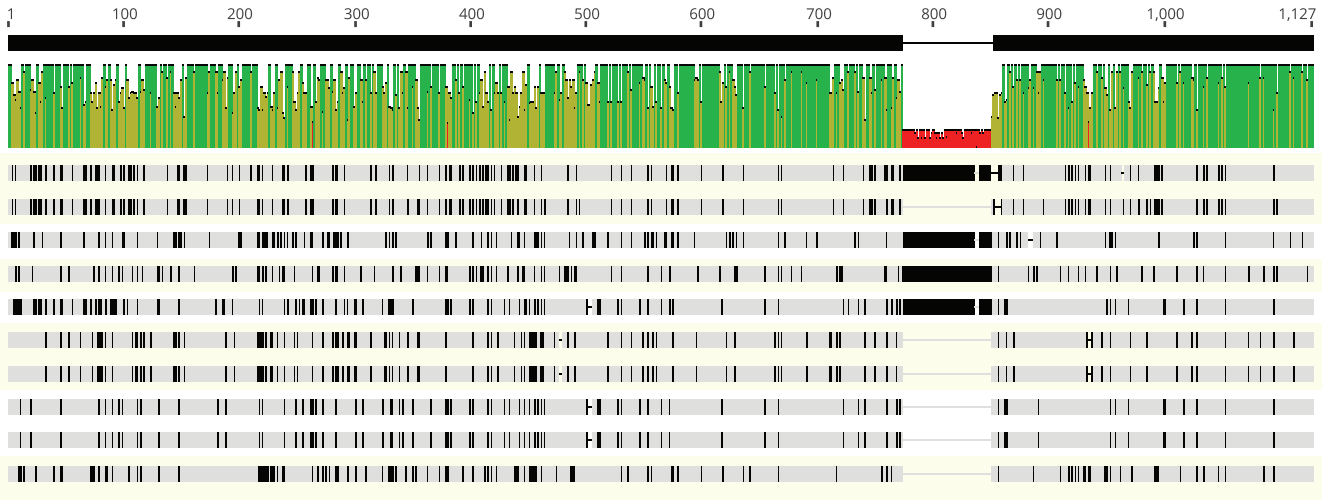
Supplementary Figure 4: Multiple sequence alignment of *de novo* reconstructed OG7530 (A) and OG9774 (B). Consensus sequences were aligned using MAFFT and visualized in Geneious Prime 2024.0.4. Each large rectangle represents an individual nucleotide sequence. Grey regions represent nucleotides identical to the consensus sequence, whereas black narrow lines represent nucleotide diversity across alleles, narrow lines represent gaps, black blocks represent inspections. Regions which allow distinguish MF-1 and LFNJM1 group (A), Au3 group (B) are highlighted in red.


A
bE3 SA
bE5 LFNJM1
bE5 MF1
bE6 MF1
bE6 LFNJM1
bE1 APG1
bE1 Au3
bE4 SA
bE2 Au3
bE7 APG1
0.980
0.792
0.950
0.957
1
0.05
B
bW5 MF1
bW7 APG1
bW5 LFNJM1
bW2 Au3
bW6 MF1
bW1* APG1
bW1 Au3
bW6 LFNJM1
bW4 SA
bW3 SA
0.986
0.987
0.959
1
0.05
C
CG-bE3
SA(ref)-bE3
LFNJM1(ref)-bE5
MF1(ref)-bE5
LFNJM1-bE5
MF1-bE5
LFNJM4-bE5
LFNEP1-bE5
LFNJRM1-bE5
3.1-bE1
LFNJM3-bE1
Au3(ref)-bE1
Au3-bE1
SYD-bE1
APG1(ref)-bE1*
CA-bE9
3.1-bE2
Au3(ref)-bE2
Au3-bE2
SYD-bE2
APG1(ref)-bE7
APG1-bE7
LFNJM3-bE7
LFENP1-bE6
LFNJM1(ref)-bE6
LFNJM1-bE6
LFNJRM1-bE6
MF1-bE6
LFNJM4-bE6
MF1(ref)-bE6
SA(ref)-bE4
CG-bE10
CA-bE8
0.882
0.443
0.994
0.737
0.336
0.574
1
0.05
D
APG1(ref)-bW1*
LFNJM3-bW1
SYD-bW1
Au3(ref)-bW1
Au3-bW1
3.1-bW1
SA(ref)-bW3
CG-bW3
LFNEP1-bW5
LFNJM1-bW5
LFNJM1(ref)-bW5
MF1-bW5
MF1(ref)-bW5
LFJRM1-bW5
LFNJM4-bW5
CG-bW10
CA-bW8
SA(ref)-bW4
MF1(ref)-bW6
LFNEP1-bW6
LFJRM1-bW6
LFJM1(ref)-bW6
MF1-bW6
LFNJM1-bW6
LFNJM4-bW6
CA-bW9
3.1-bW2
SYD-bW2
Au3(ref)-bW2
Au3-bW2
APG1(ref)-bW7
LFNJM3-bW7
0.512
0.41
0.594
0.999
0.994
1
0.05
bW1/bE1
bW2/bE2
Pandemic lineage
Pimenta dioica
bW5/bE5
bW6/bE6
Eucalyptus sp.
Syzygium sp.
bW1/bE1
bW7/bE7
Psidium guajava
Syzygium sp.
Eugenia stipitata
bW3/bE3
bW4/bE4
South Africa (SA) lineage
bW3/bE3
bW10/bE10
Eugenia dysentericus
bW8/bE8
bW9/bE9
Plinia edulis

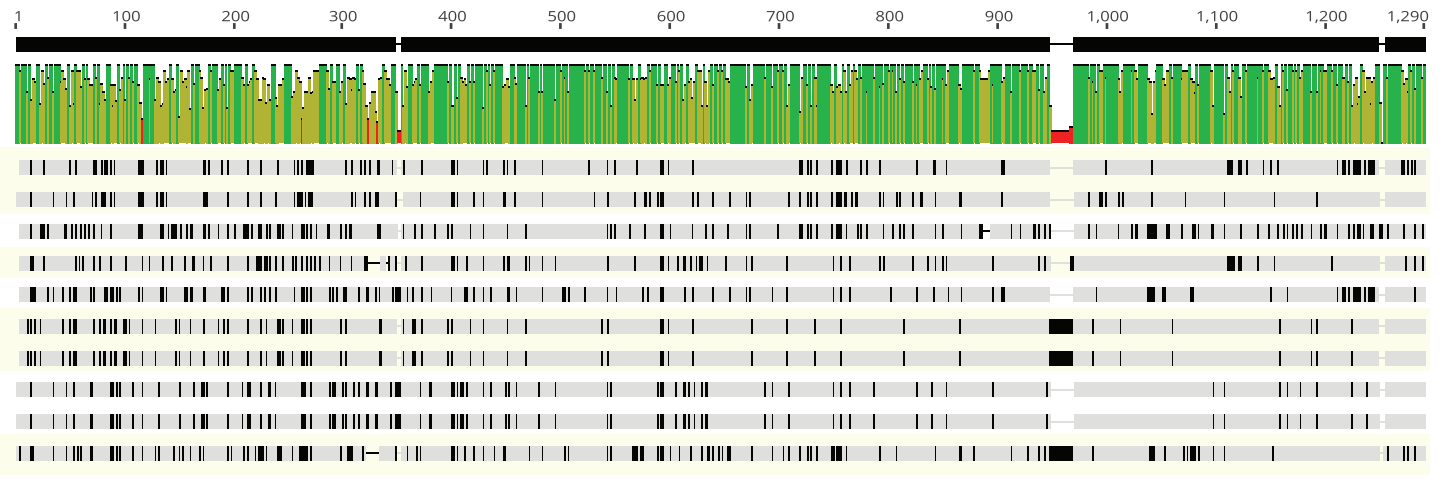

A
APG1
APG1(ref)
LFNJM3
CA
3.1
CG
LFNJRM1
LFNJM4
LFNEP1
LFNJM1
LFNJM1(ref)
MF-1(ref)
MF-1
Au3
Au3(ref)
SYD
PC 2
PC 1
15
10
5
0
−5
−20
−15
−10
−5
0
5


B
APG1
APG1(ref)
LFNJM3
CA
LFNJRM1
CG
3.1
LFNJM4
Au3(ref)
LFNEP1
LFNJM1
Au3
LFNJM1(ref)
SYD
MF-1
MF-1(ref)
PC 2
PC 1
15
10
5
0
−5
−5
0
5
10
15
20

A


1
100
200
300
400
500
600
700
800
900
1,000
1,127
Consensus Sequence
bE1-HD2 Au3(ref)
bE1-HD2 APG1(ref)
bE2-HD2 Au3(ref)
bE3-HD2 SA(ref)
bE4-HD2 SA(ref)
bE5-HD2 MF1ref)
bE5-HD2 JM1ref)
bE6-HD2 JM1(ref)
bE6-HD2 MF1(ref)
bE7-HD2 APG1ref)

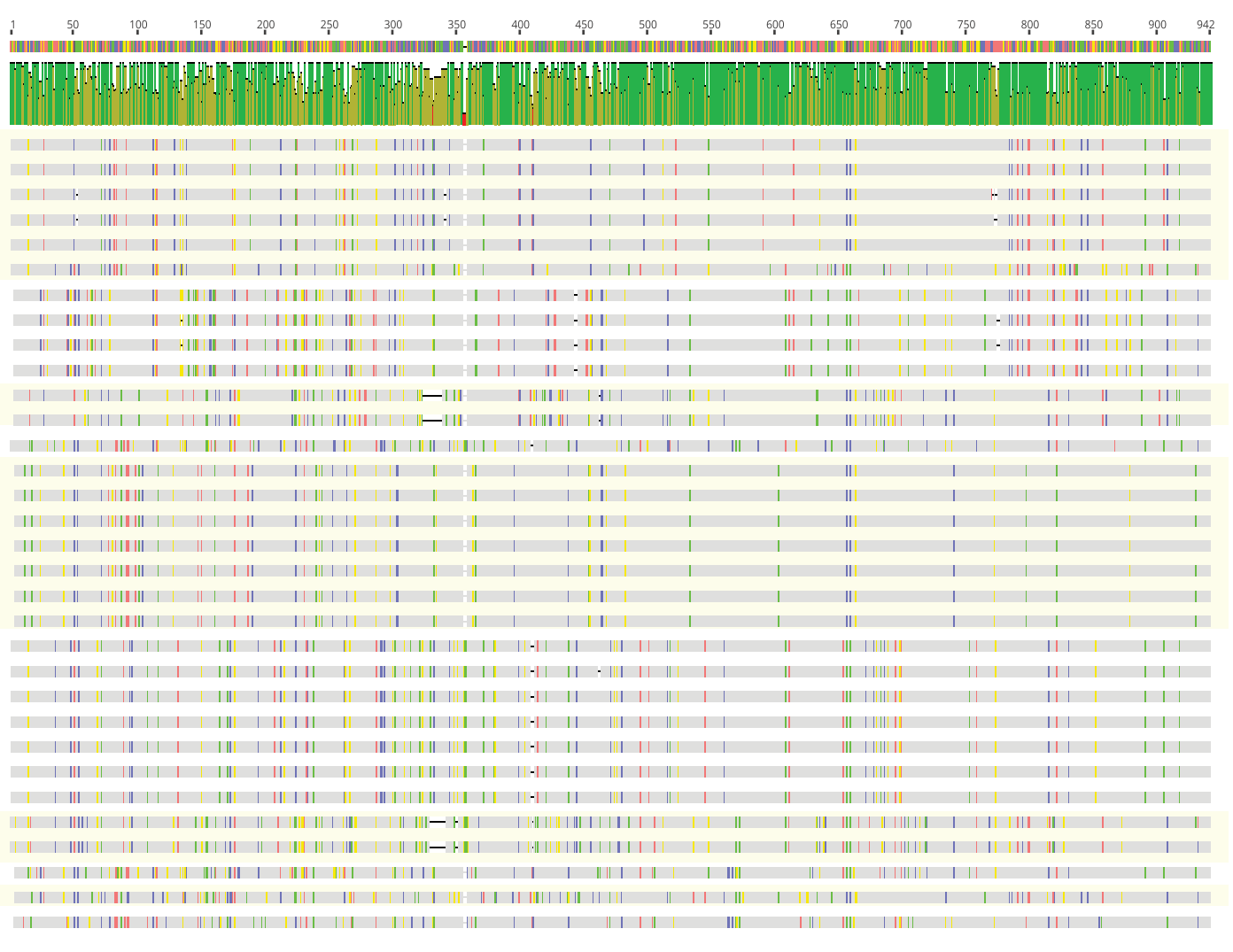
B

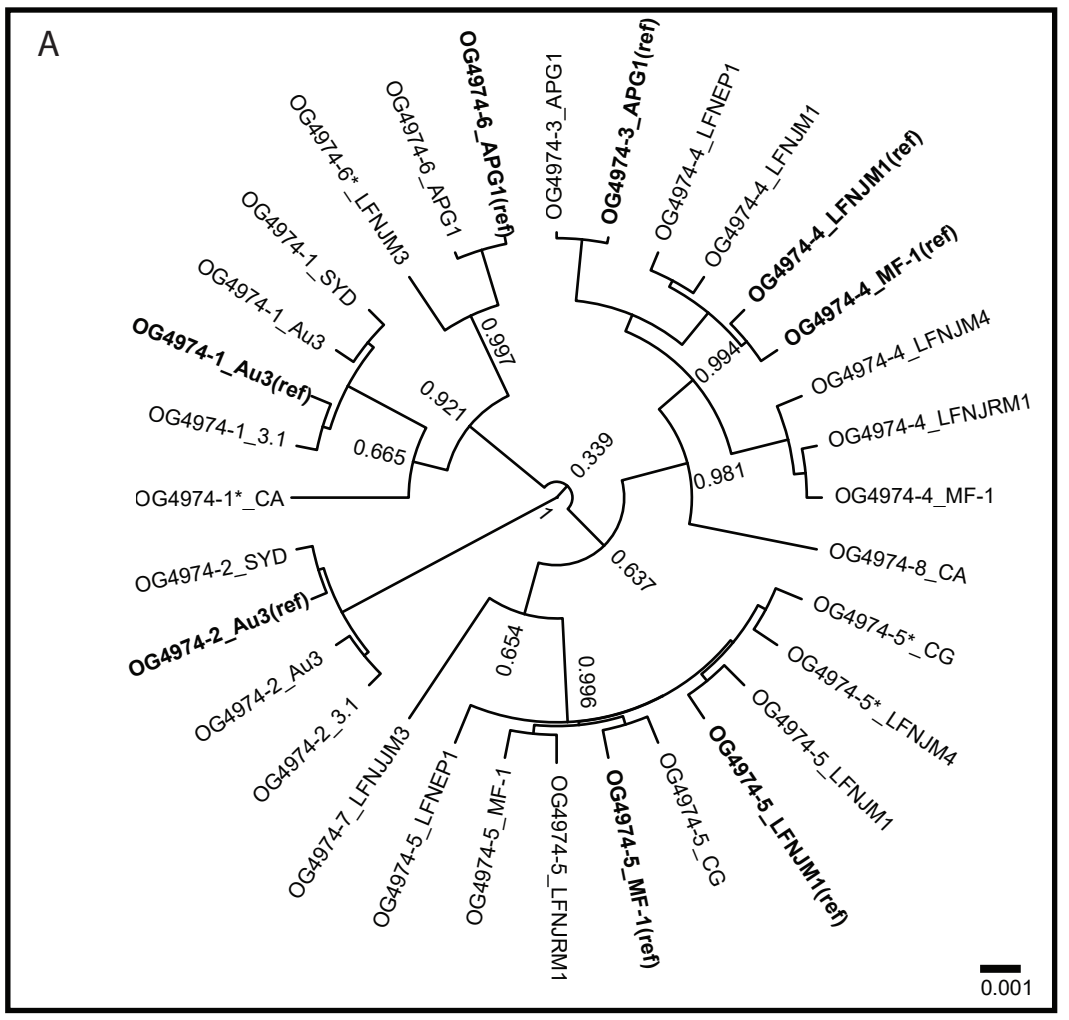

1
100
200
300
400
500
600
700
800
900
1,000
1,100
1,200
1,290
Consensus Sequence
bW1-HD2 Au3(ref)
bW1-HD2* APG1(ref)
bW2-HD2 Au3(ref)
bW3-HD2 SA(ref)
bW4-HD2 SA(ref)
bW5-HD2 MF1ref)
bW5-HD2 JM1ref)
bW6-HD2 JM1(ref)
bW6-HD2 MF1(ref)
bW7-HD2 APG1ref)

C

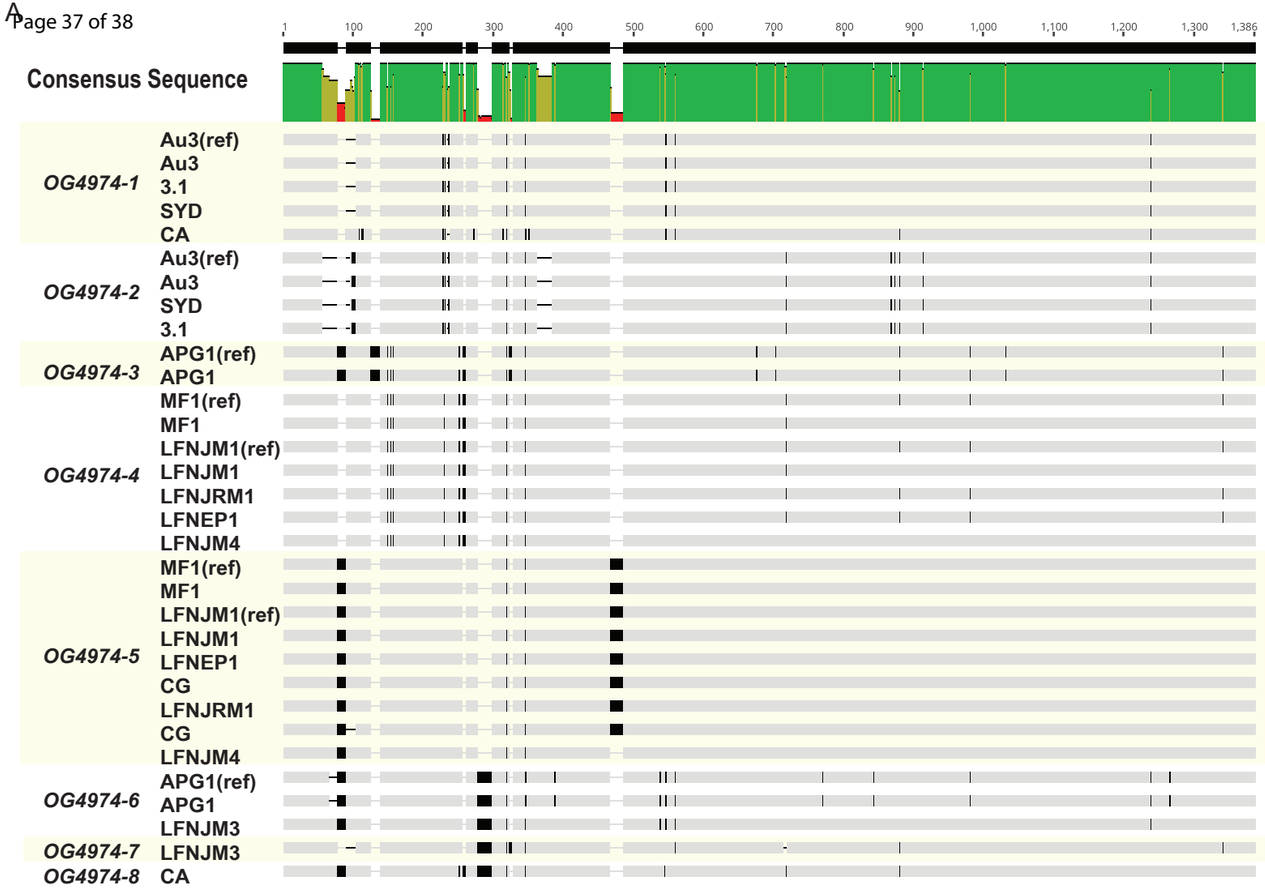

1
50
100
150
200
250
300
350
400
450
500
550
600
657
Consensus Sequence
bE1-HD2 Au3(ref)
Au3
3.1
SYD
LFNJM3
APG1(ref)
bE2-HD2 Au3(ref)
Au3
SYD
3.1
bE3-HD2 SA
CG
bE4-HD2 SA
bE5-HD2 MF1(ref)
MF1
LFNJM1(ref)
LFNJM1
LFNJRM1
LFNEP1
LFNJM4
bE6-HD2 MF1(ref)
MF1
LFNJM1(ref)
LFNJM1
LFNJRM1
LFNEP1
LFNJM4
bE7-HD2 APG1(ref)
APG1
LFNJM3
bE8-HD2 CA
bE9-HD2 CA
bE10-HD2 CG

D

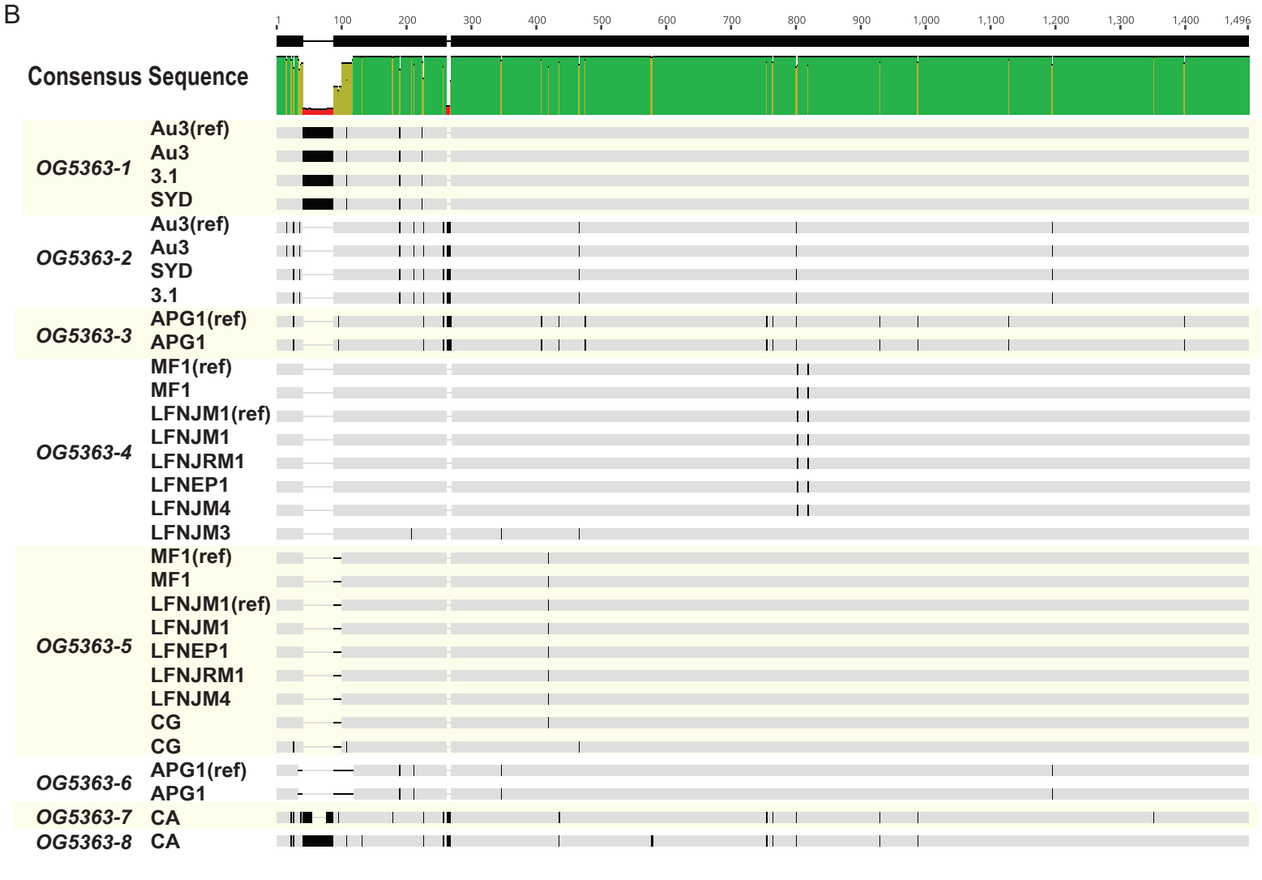

1
50
100
150
200
250
300
350
400
450
500
550
600
650
700
750
800
850
900
942
Consensus Sequence
bW1-HD1 Au3(ref)
Au3
3.1
SYD
LFNJM3
APG1(ref)
bW2-HD1 Au3(ref)
Au3
SYD
3.1
bW3-HD1 SA
CG
bW4-HD1 SA
bW5-HD1 MF1(ref)
MF1
LFNJM1(ref)
LFNJM1
LFNJRM1
LFNEP1
LFNJM4
bW6-HD1 MF1(ref)
MF1
LFNJM1(ref)
LFNJM1
LFNJRM1
LFNEP1
LFNJM4
bW7-HD1 APG1(ref)
LFNJM3
bW8-HD1 CA
bW9-HD1 CA
bW10-HD1 CG


A
OG4974-6_APG1(ref)
OG4974-6_APG1
OG4974-6*_LFNJM3
OG4974-1_SYD
OG4974-1_Au3
OG4974-1_Au3(ref)
OG4974-1_3.1
OG4974-1*_CA
OG4974-2_SYD
OG4974-2_Au3(ref)
OG4974-2_Au3
OG4974-2_3.1
OG4974-7_LFNJJM3
OG4974-5_LFNEP1
OG4974-5_MF-1
OG4974-5_LFNJRM1
OG4974-5_MF-1(ref)
OG4974-5_CG
OG4974-5_LFNJM1
OG4974-5*_LFNJM4
OG4974-5*_CG
OG4974-8_CA
OG4974-4_MF-1
OG4974-4_LFNJRM1
OG4974-4_LFNJM4
OG4974-4_MF-1(ref)
OG4974-4_LFNJM1(ref)
OG4974-4_LFNJM1
OG4974-4_LFNEP1
OG4974-3_APG1(ref)
OG4974-3_APG1
0.997
0.921
0.665
1
0.339
0.994
0.981
0.637
0.654
0.996
0.001
B
OG5363-5_MF-1(ref)
OG5363-5_LFNJM4
OG5363-5_LFNJRM1
OG5363-5_CG
OG5363-5_MF-1
OG5363-5_LFNJM1(ref)
OG5363-5_LFNJM1
OG5363-5_LFNEP1
OG5363-4_MF-1
OG5363-4_LFNJM1(ref)
OG5363-4_LFNJRM1
OG5363-4_LFNJM1
OG5363-4_MF-1(ref)
OG5363-4_LFNJM4
OG5363-4_LFNEP1
OG5363-5*_CG
OG5363-4*_LFNJM3
OG5363-6_APG1(ref)
OG5363-6_APG1
OG5363-1_SYD
OG5363-1_Au3(ref)
OG5363-1_Au3
OG5363-1_3.1
OG5363-8_CA
OG5363-7_CA
OG5363-3_APG1(ref)
OG5363-3_APG1
OG5363-2_Au3(ref)
OG5363-2_Au3
OG5363-2_SYD
OG5363-2_3.1
0.854
1
0.994
0.397
0.958
0.473
0.449
0.001
C
OG7071-4_MF-1(ref)
OG7071-4_MF-1
OG7071-4_LFNJRM1
OG7071-4_LFNJM4
OG7071-4_LFNJM1
OG7071-4_LFNJM1(ref)
OG7071-4_LFNEP1
OG7071-3_APG1(ref)
OG7071-3_APG1
OG7071-6_LFNJM3
OG7071-6_LFNJM3
OG7071-5_LFNJM1(ref)
OG7071-5_LFNJM1
OG7071-5_MF-1
OG7071-5_CG
OG7071-5_MF-1(ref)
OG7071-5_LFNJRM1
OG7071-5_LFNJM4
OG7071-5_LFNEP1
OG7071-1_SYD
OG7071-1_Au3(ref)
OG7071-1_Au3
OG7071-1_3.1
OG7071-1_APG1(ref)
OG7071-1_APG1
OG7071-7_CA
OG7071-2*_CA
OG7071-2_3.1
OG7071-2_Au3(ref)
OG7071-2_Au3
OG7071-2_Syd
1
1
1
0.953
1
0.704
0.793
1
1
0.001
D
OG7530-6_MF-1(ref)
OG7530-6_LFNJM1
OG7530-6_LFNJRM1
OG7530-6_LFNJM1(ref)
OG7530-6_LFNJM4
OG7530-6_LFNEP1
OG7530-7_CA
OG7530-3_APG1(ref)
OG7530-3_APG1
OG7530-4_LFNJM3
OG7530-4_APG1(ref)
OG7530-4_APG1
OG7530-2_LFNJM3
OG7530-2_CA
OG7530-2_SYD
OG7530-2_Au3
OG7530-2_Au3(ref)
OG7530-2_3.1
OG7530-2_CG
OG7530-1_SYD
OG7530-1_Au3(ref)
OG7530-1_Au3
OG7530-1_CA
OG7530-1_3.1
OG7530-5_MF-1(ref)
OG7530-5_LFNJM4
OG7530-5_MF-1
OG7530-5_LFNEP1
OG7530-5*_LFNJM1(ref)
OG7530-5*_LFNJRM1
OG7530-5*_LFNJM1
1
1
1
0.370
0.439
0.337
0.654
0.995
0.828
0.998
0.001
E
OG9632-2_Au3(ref)
OG9632-2_Au3
OG9632-2_SYD
OG9632-2_3.1
OG9632-7_APG1(ref)
OG9632-7_GM1
OG9632-6_CA
OG9632-5_APG1(ref)
OG9632-5_APG1
OG9632-3_LFNJM4
OG9632-3_LFNJM1(ref)
OG9632-3_MF-1(ref)
OG9632-3_LFNJM1
OG9632-3_MF-1
OG9632-3_LFNJRM1
OG9632-3_LFNEP1
OG9632-1_SYD
OG9632-1_Au3(ref)
OG9632-1_Au3
OG9632-1_3.1
OG9632-1_LFNJM3
OG9632-9_CA
OG9632-8_LFNJM3
OG9632-8_CG
OG9632-4_MF-1(ref)
OG9632-4_MF-1
OG9632-4_LFNJM1(ref)
OG9632-4_LFNJRM1
OG9632-4_LFNJM1
OG9632-4_LFNJM4
OG9632-4_LFNEP1
OG9632-4_CG
0.454
1
0.570
0.823
1
0.706
0.986
0.553
0.981
0.001
F
OG9774-4_MF-1(ref)
OG9774-4_LFNJRM1
OG9774-4_LFNJM1
OG9774-4_MF-1
OG9774-4_CA
OG9774-5_APG1
OG9774-5_APG1(ref)
OG9774-5_LFNJM3
OG9774-4_LFNJM4
OG9774-4_LFNJM1(ref)
OG9774-4_LFNEP1
OG9774-4_CG
OG9774-4_CG
OG9774-6_APG1
OG9774-6_APG1(ref)
OG9774-6_LFNJM3
OG9774-1*_3.1
OG9774-1_Au3
OG9774-1_Au3(ref)
OG9774-1*_CA
OG9774-1_SYD
OG9774-3_LFNJM4
OG9774-3_MF-1(ref)
OG9774-3_LFNJM1(ref)
OG9774-3_MF-1
OG9774-3_LFNJM1
OG9774-3_LFNJRM1
OG9774-3_LFNEP1
OG9774-2_Au3(ref)
OG9774-2_Au3
OG9774-2_SYD
OG9774-2_3.1
1
1
1
0.575
0.362
0.508
0.418
0.083
0.999
0.954
0.989
0.001

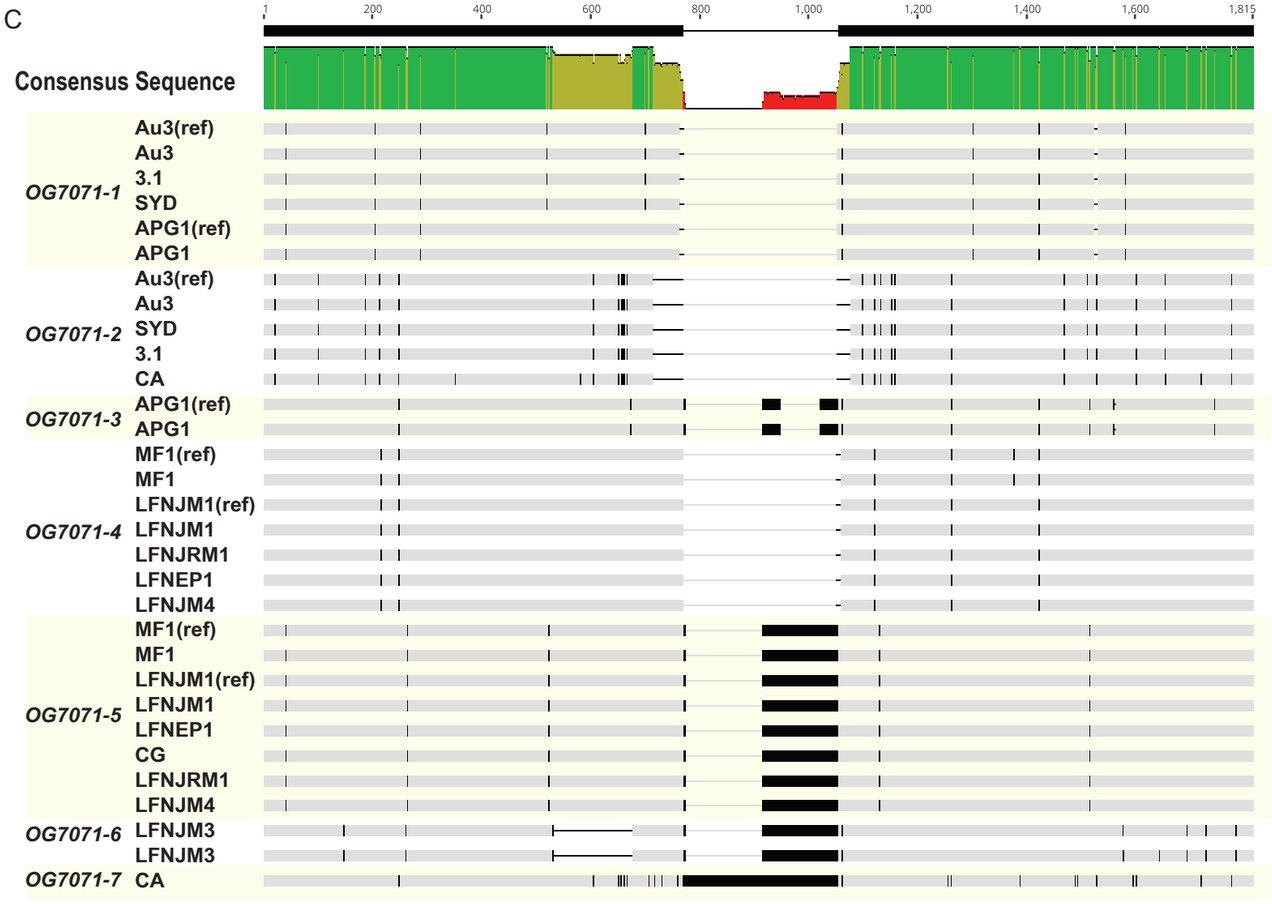
A


Consensus Sequence
1 100 200 300 400 500 600 700 800 900 1,000 1,100 1,200 1,300 1,386
OG4974-1 Au3(ref) Au3 3.1 SYD CA
OG4974-2 Au3(ref) Au3 SYD 3.1
OG4974-3 APG1(ref) APG1
OG4974-4 MF1(ref) MF1 LFNJM1(ref) LFNJM1 LFNJRM1 LFNEP1 LFNJM4
OG4974-5 MF1(ref) MF1 LFNJM1(ref) LFNJM1 LFNEP1 CG LFNJRM1 CG LFNJM4
OG4974-6 APG1(ref) APG1 LFNJM3
OG4974-7 LFNJM3
OG4974-8 CA

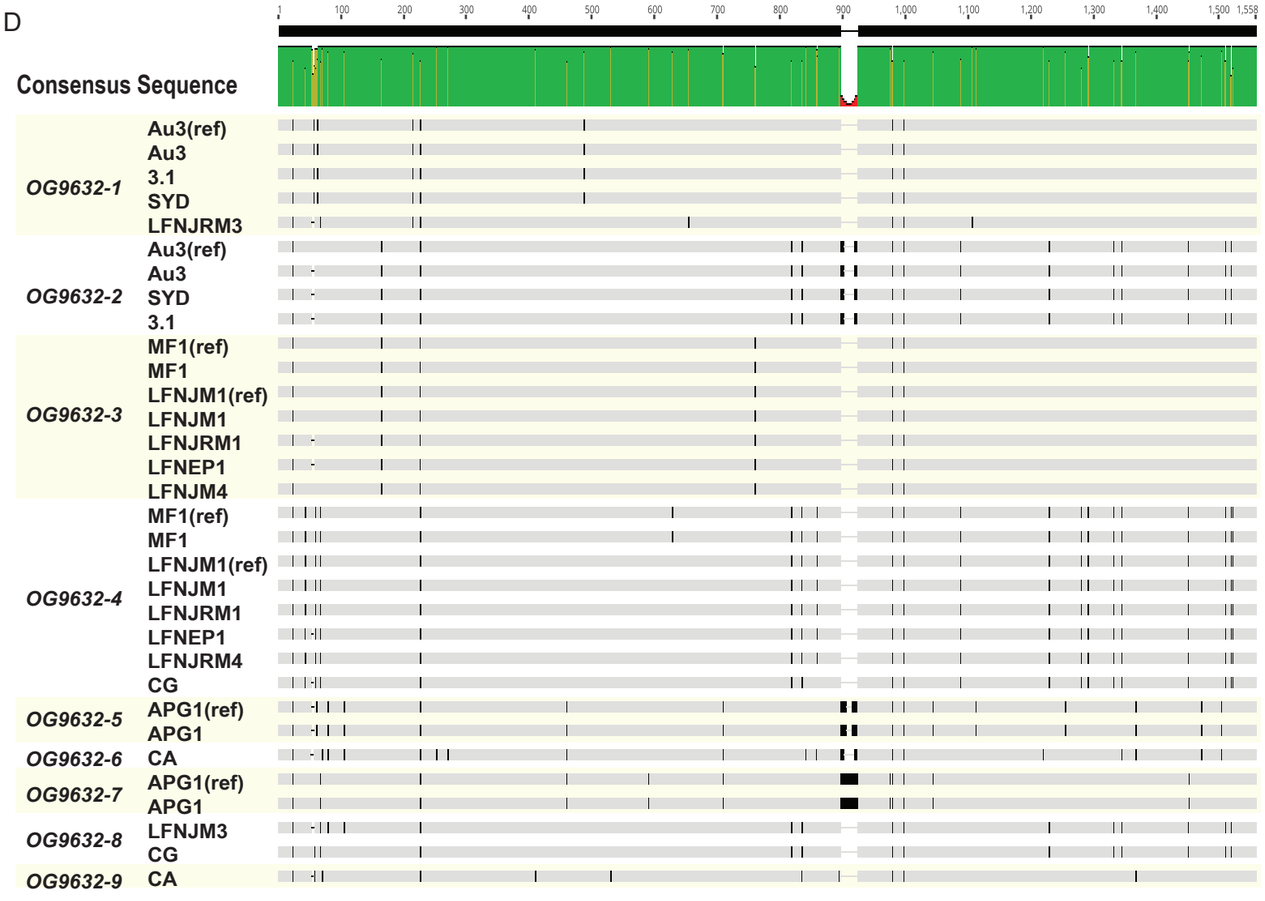
B


Consensus Sequence
1 100 200 300 400 500 600 700 800 900 1,000 1,100 1,200 1,300 1,400 1,496
OG5363-1 Au3(ref) Au3 3.1 SYD
OG5363-2 Au3(ref) Au3 SYD 3.1
OG5363-3 APG1(ref) APG1
OG5363-4 MF1(ref) MF1 LFNJM1(ref) LFNJM1 LFNJRM1 LFNEP1 LFNJM4 LFNJM3
OG5363-5 MF1(ref) MF1 LFNJM1(ref) LFNJM1 LFNEP1 LFNJRM1 LFNJM4 CG CG
OG5363-6 APG1(ref) APG1
OG5363-7 CA
OG5363-8 CA

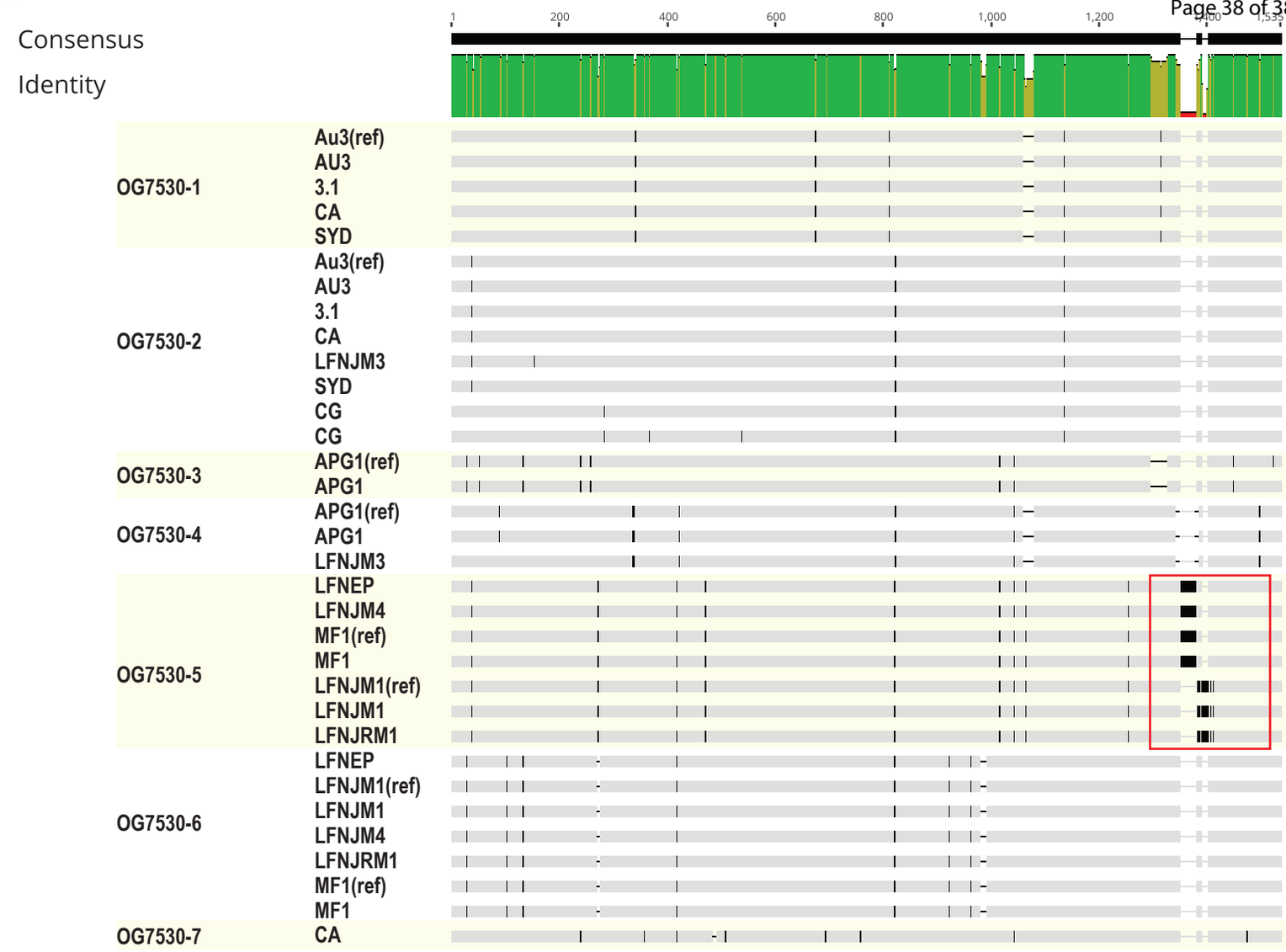
C


Consensus Sequence
1 200 400 600 800 1,000 1,200 1,400 1,600 1,815
OG7071-1 Au3(ref) Au3 3.1 SYD APG1(ref) APG1
OG7071-2 Au3(ref) Au3 SYD 3.1 CA
OG7071-3 APG1(ref) APG1
OG7071-4 MF1(ref) MF1 LFNJM1(ref) LFNJM1 LFNJRM1 LFNEP1 LFNJM4
OG7071-5 MF1(ref) MF1 LFNJM1(ref) LFNJM1 LFNEP1 CG LFNJRM1 LFNJM4
OG7071-6 LFNJM3 LFNJM3
OG7071-7 CA

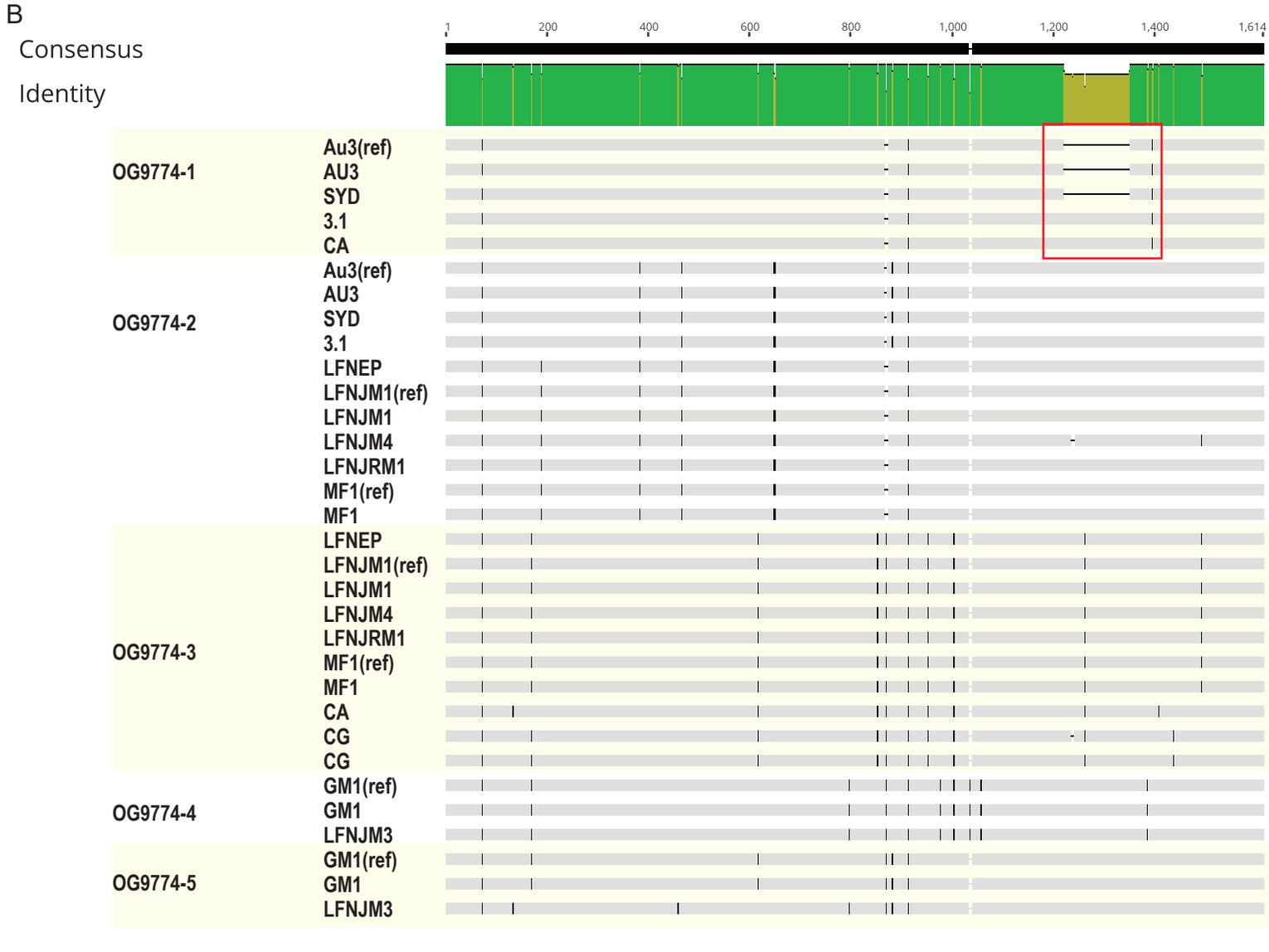
D


Consensus Sequence
1 100 200 300 400 500 600 700 800 900 1,000 1,100 1,200 1,300 1,400 1,500 1,558
OG9632-1 Au3(ref) Au3 3.1 SYD LFNJRM3
OG9632-2 Au3(ref) Au3 SYD 3.1
OG9632-3 MF1(ref) MF1 LFNJM1(ref) LFNJM1 LFNJRM1 LFNEP1 LFNJM4
OG9632-4 MF1(ref) MF1 LFNJM1(ref) LFNJM1 LFNJRM1 LFNEP1 LFNJRM4 CG
OG9632-5 APG1(ref) APG1
OG9632-6 CA
OG9632-7 APG1(ref) APG1
OG9632-8 LFNJM3 CG
OG9632-9 CA

A


Consensus
Identity
1
200
400
600
800
1,000
1,200
1,400
1,535
OG7530-1
Au3(ref)
AU3
3.1
CA
SYD
OG7530-2
Au3(ref)
AU3
3.1
CA
LFNJM3
SYD
CG
CG
OG7530-3
APG1(ref)
APG1
OG7530-4
APG1(ref)
APG1
LFNJM3
OG7530-5
LFNEP
LFNJM4
MF1(ref)
MF1
LFNJM1(ref)
LFNJM1
LFNJRM1
OG7530-6
LFNEP
LFNJM1(ref)
LFNJM1
LFNJM4
LFNJRM1
MF1(ref)
MF1
OG7530-7
CA

B


Consensus
Identity
1
200
400
600
800
1,000
1,200
1,400
1,614
OG9774-1
Au3(ref)
AU3
SYD
3.1
CA
OG9774-2
Au3(ref)
AU3
SYD
3.1
LFNEP
LFNJM1(ref)
LFNJM1
LFNJM4
LFNJRM1
MF1(ref)
MF1
OG9774-3
LFNEP
LFNJM1(ref)
LFNJM1
LFNJM4
LFNJRM1
MF1(ref)
MF1
CA
CG
CG
OG9774-4
GM1(ref)
GM1
LFNJM3
OG9774-5
GM1(ref)
GM1
LFNJM3