From populations to pan-genomes: investigating the role of ecology and genomic architecture in maintaining species boundaries in the porcini mushroom, *Boletus edulis*

Keaton Tremble[1,4], Etienne Brejon Lamartinière[2], Alexander J. Bradshaw[1,4], Roberto Flores Arzú[3], Joseph Hoffman[2], and Bryn T.M. Dentinger[1,4]

[1]School of Biological Sciences, University of Utah, Salt Lake City, USA
[2]Department of Animal Behaviour, Bielefeld University, Bielefeld, Germany
[3]Facultad de Ciencias Químicas y Farmacia, Universidad de San Carlos de Guatemala, Ciudad de Guatemala, Guatemala
[4]Natural History Museum of Utah, Salt Lake City, USA

September 5, 2023

Abstract

The forces that maintain stable boundaries between closely related groups are poorly characterized in the *Fungi*. In other organisms prezygotic barriers, such as assortative mating, are sufficient to reinforce local adaptation and prevent hybridization, yet assortative mating and other prezygotic barriers are largely absent in most fungi. Postzygotic barriers, such as large structural changes across the genome, can also maintain lineage segregation through hybrid incompatability. The role of structural variation in speciation has been well characterized in a wide array of animals and plants, but not in the *Fungi*, where it may be relatively uncommon. Here, we use a pan-genomic approach in the globally distributed ectomycorrhizal fungus *Boletus edulis* Bull. with 310 whole genome sequences and 22 newly generated reference genomes to identify whether large genomic changes are necessary to maintain lineage segregation in the presence of ongoing gene flow. In addition, we evaluate the extent to which different lineages possess unique repertories of saprotrophic and effector gene families to determine whether lineages exhibit distinct ecological capacities that contribute toward lineage segregation. We show that *B. edulis* possesses one of the most diverse pan-genomes of eukaryotes, with the lowest proportion of core proteins identified in the *Fungi*. Pan-genomic

clustering correlates more with geography than with gene flow or phylogenetic relatedness. We also show that structural variation is abundant within *B. edulis*, but is rarely fixed within a lineage or even shared between two individuals. Moreover, we find that a similar degree of genomic synteny exists between genomes of different lineages as between genomes of the same lineage. Functional protein analysis shows that a new lineage of *B. edulis* from tropical Guatemala has significant contractions in most saprotrophic gene families and possesses a distinct saprotrophic enzyme repretoire. This lineage also a carries a clear signial of a demographic bottleneck, has very low genome-wide heterozygosity, and an inflated genome size, consistent with a substantial shift in ecological niche. In addition, we reveal preliminary evidence for subtle differences in host preference among European *B. edulis*. Altogether, our results provide clear evidence that large changes in genome structure or proteome content are not necessary or perhaps even sufficient to maintain long-term lineage fidelity in the presence of ongoing gene flow. Instead, it appears that subtle differences in ecological tendencies can maintain stable boundaries through time.

1 Introduction

Identifying the evolutionary forces that contribute to stable species boundaries has been one of the fundamental pursuits of evolutionary biology since the modern synthesis. However, these forces remain poorly characterized in the vast majority of life. The genomic revolution has provided researchers innumerable insights into the population processes that initiate genomic divergence between groups, but the presence of divergence alone is not necessarily indicative of incipient speciation. For example, population divergence can develop rapidly during climatic upheaval, only to erode when gene-flow is resumed or distinct habitats become homogenized (Ravinet et al. 2021). For divergent lineages to remain divergent and to move towards speciation, evolutionary forces must act on existing genetic diversity to maintain segregation through time. For many animals (Maan et al. 2011; West-Eberhard 1983), and some plants (Moore et al. 2011), sexual selection and other forms of assortative mating are sufficient to reinforce reproductive barriers, generating stable lineage boundaries. However, in groups without complex assortative mating, such as *Fungi*, it remains unclear what evolutionary forces are necessary or sufficient to maintain lineage segregation. Population genomic surveys of diverse fungi, such as *Neurospera crassa* Shear & B.O. Dodge (Ellison et al. 2011), *Suillus brevipes* (Peck) Kuntze (Branco et al. 2017), and *Boletus edulis* Bull. (Tremble, Hoffman, et al. 2022), have demonstrated that adaptation to local environmental conditions is likely the primary driver of initial genomic divergence. Yet, in each of these systems strong signatures of gene flow remain between divergent groups post adaptation. Are the processes that generate genomic diversity, such as local adaptation, sufficient to maintain divergence over time, or are additional reinforcing genomic conflicts, such as proteome turnover and chromosomal re-arrangements, necessary for speciation?

The role that genomic structural variation plays in the process of evolutionary divergence is paradoxically well characterized but poorly understood. Large chromosomal

inversions, translocations, and abundant insertions and deletions can lead to hydrid dysfunction, where rearrangements result in the incorrect recombination of chromosomes and the formation of underdominant gametes (Ayala et al. 2005; L. Zhang et al. 2021). Structural rearrangements can also facilitate genomic divergence by suppressing recombination between heterozygous chromosomes, thereby sheltering locally adaptive alleles from the homogenizing effects of gene flow (Navarro et al. 1997; Rieseberg 2001). Lastly, insertions and deletions can directly influence patterns of divergence if the genomic regions involved contain gene(s) that encode advantageous phenotype(s), or that disrupt the expression of other loci (Hämälä et al. 2021). These mechanisms have been well characterized in many animal and plant species, yet are poorly understood in *Fungi*, where they may be relatively uncommon. It has been suggested that true genomic incompatibilities are slow to form, or rarely form at all, in the *fungi* (Giraud et al. 2012; Gladieux et al. 2014). As evidence, interspecific hybridization that leads to viable and often advantageous offspring, would appear to be at least as common and widespread among *Fungi* as plants, and far more common than animals (Feurtey et al. 2018; Steensels et al. 2021). Furthermore, it has been demonstrated that hybridization of heteroploid parental species can lead to sexually viable aneuploid hybrids (Inderbitzin et al. 2011; Shoji et al. 2015), and that the distinct lack of intraspecific genomic synteny, termed mesosyteny, is a diagnosable characteristic of both sexual and asexual *Dothidiomycetes* (Hane et al. 2011). This suggests that structural variation, which would be expected to produce hybrid dysfunction, is less consequential in *Fungi* than in other eukaryotes. Yet, structural variation still likely causes hybrid sterility in some groups, such as hybrids between *Cryptococcus neoformans* and *C. deneoformans* (Samarasinghe et al. 2020). It is therefore unclear to what extent large structural variations affect the speciation process in the *Fungi*, and to what extent structural variation can persist within a single evolutionary unit.

One approach to identify the extent of genome-wide variation, both structural and genetic, is to characterize the pan-genome of a species. Pan-genomic approaches use *de novo* genome assemblies to identify the gene families that are "core"— families that are present in all genomes, and are likely necessary for key molecular and ecological functions of the species — and "accessory" – families that are present in some but not all genomes, and are potentially involved in local adaptation. Population genomic analyses of genome-wide variation have produced innumerable insights into the process of speciation (Seehausen et al. 2014). However, these approaches rely on mapping sequences to a single reference genome, and fail to capture variation present in the population that is not in the reference. Pan-genomic analyses capture the total genetic potential of a species, and have revealed an enormous amount of genetic diversity and structural variation that was previously unidentifiable (Hirsch et al. 2014; Y. Liu et al. 2020; Sherman et al. 2020; Vernikos 2020). Not only does this approach characterize the proteome of a specimen, it is also one measure of structural variation that exists between two genomes, as accessory proteins are often the product of insertions, deletions, and duplication events (L. Zhang et al. 2021). To characterize the role that pan-genomic diversity and structural variation play in the fungal speciation process, we therefore reconstruct the first global pan-genome of an ectomycorrhizal fungus (EMF).

The ecological forces that contribute to fungal divergence are poorly characterized, particularly within the EMF, a hyper-diverse group of soil-borne filamentous fungi that form mutualisistic relationships with plants. EMF possess a cryptic lifestyle that has remained frustratingly difficult to elucidate, yet several patterns of the EMF "syndrome" emerged and are supported by a general consensus. On the other hand, some ecological functions of EMF remain contentious. For instance, it is well accepted that EMF play an integral role in soil carbon cycling (Lindahl et al. 2015; Steidinger et al. 2019; J. M. Talbot et al. 2008). However, it remains a matter of debate whether they only decompose organic material to "mine" nutrients for provision to plant hosts, or whether some EMF are "facultative-saprotrophs" and can acquire carbon through both decomposition and mycorrhizal mutualsm (Baldrian 2009; Jennifer M. Talbot et al. 2013; Vaario et al. 2019). Regardless, it is evident that decomposition is an important component of the EMF lifestyle (Kohler et al. 2015; Jonathan M Plett et al. 2023) and variation in saprotrophic capacity may contribute to lineage divergence.

Host preference and cryptic patterns of host specialization may also facilitate EMF divergence. While the genetic toolkit used by EMF to colonize a host is altogether poorly understood, much work has shown that small secreted effector proteins that localize to the host nucleus are necessary for colonization in the ectomycorrhizal fungus *Laccaria bicolor* (Maire) P.D. Orton (Kang et al. 2020; Pellegrin et al. 2019). However, EMF effector proteins diverge rapidly and are often species specific, which yields few consistent patterns of effector protein abundance and degree of host specialization or generalism (Lofgren, N. H. Nguyen, et al. 2021). Futhermore, host specialization at the level of species is considered scarce, with host associations thought to be more important at broader evolutionary scales (Ishida et al. 2007). Lastly, within the genome, much recent work has documented patterns of that transposable elements (TEs) abundance that may play a role in the adaptive evolution of EMF (Jonathan M Plett et al. 2023; Wu et al. 2022). In particular, most EMF exhibit expanded TEs families and inflated genome sizes compared to their saprotrophic ancestors, and these TEs are often clustered around rapidly diverging genes such as effector proteins (Jonathan M Plett et al. 2023; Wu et al. 2022).

The prized edible porcini mushroom, *Boletus edulis*, is a globally distributed ectomycorrhizal fungus with an exceptional ecological amplitude. *B. edulis* can be found in nearly every temperate ecosystem in the northern hemisphere; from the tropics in Guatemala, to north of the arctic circle in Alaska (Tremble, Suz, et al. 2020). Despite being nearly continuously distributed across this vast geographic range wherever its plant symbionts occur, previous work has shown that *B. edulis* shows strong cryptic, geographically distinct population structure, with limited admixture (Tremble, Hoffman, et al. 2022). Demographic and ecological niche modeling suggest that these lineages diverged over the past 2.6 million years in the presence of continuous gene flow, and that they exhibit little to no within-lineage population structure. This supports the hypothesis that local adaptation is the primary force generating lineage diversity within *B. edulis*. However, despite clear population structure and high levels of genomic divergence, there are no consistent patterns of ecological, phe-

nological, or even morphological differentiation among these lineages. This suggests that, if phenotypic differences do exist, these differences are cryptic, or that the conditions that initially generated them have since changed.

Here, we evaluate whether lineage segregation within *B. edulis* is maintained by large genomic structural variation reinforcing local adaptation through hybrid incompatibility, or by smaller changes in the abundance of specific ecologically relevant gene families acted upon by natural selection. First, we generated 310 whole genome sequences, 21 new reference-quality genome assemblies using a hybrid sequencing approach, an additional chromosomal level reference assembly, and assigned these genomes to lineages using a summary-coalescent phylogenomic approach. After lineage placement, we then calculated admixture between lineages and used a coalescent method to estimate historic changes in the effective population size of each lineage. To identify broad changes in genome content and structure, we used these genome assemblies to characterize the pan-genome of *B. edulis* and clustered specimens based on pan-genomic similarity. We then performed whole-genome mapping of our 21 reference-quality hybrid genome assemblies against our chromosomal-level assembly to predict large translocations and inversions. After prediction, we characterized the extent to which these structural variants are shared within and between lineages, and investigated whether structural variants are diverging faster than the genome background. Finally, in order to identify changes in the abundance of ecologically relevant gene families we predicted carbohydrate-active enzymes (CAZymes), extracellular proteases, extracellular effector proteins, and transposable elements within all 22 of our reference genomes (21 hybrid reference genomes plus 1 chromosomal level assembly). We hypothesized that (i) structural variation is rampant within *B. edulis* but is rarely lineage-specific, (ii) the pan-genome of *B. edulis* is exceptionally diverse, but that patterns of protein presence/absence largely correspond with phylogeny, and (iii) the saprotrophic and effector protein profile of specimens within *B. edulis* is highly variable, and no clear patterns of lineage-specific divergence exist.

We report a striking diversity of genomic structural variation within *B. edulis*, including one of the most diverse pan-genomes identified within all eukaryotes and the most diverse yet identified in the fungi. Clustering of *B. edulis* specimens based on pan-genomic profiles revealed four distinct pan-genomic groups that surprisingly did not correspond with phylogeny or gene flow, suggesting that local geography is important in shaping the total genomic repertoire of a species. Large structural variation is common, yet rarely shared between two genomes, and when shared, it is more often shared between genomes of separate lineages than genomes of the same lineage. We further show that while structural variants can be "hotspots" of genomic divergence, they are likely not the cause of differentiation, but rather facilitate divergence through sheltering locally adaptive loci. Importantly, we demonstrate that the saprotrophic gene repertoire of *B. edulis* is highly variable, with significant differences being present among lineages. Specifically we show that the tropical lineage of *B. edulis* from Guatemala exhibits significant contractions in most saprotrophic gene families, and possesses a distinct saprotrophic repertoire suggesting a greater reliance on host-derived carbon sources. In addition, we identify significant differences among lin-

eages in the abundance of cytoplasmic effector proteins, yet none of the lineages possess lineage-specific effector proteins. This suggests that if cryptic differences in host preference exist among lineages, these are differences are likely linked to colonization preference rather than compatibility. Taken together, our findings suggest that long-term lineages boundaries in *Fungi* are not maintained by large changes in genome content and structure, but rather by subtle differences in ecologically relevant gene families implicating differences in ecological lifestyle as a primary driver of speciation in fungi.

2 Methods

2.1 Specimen Collection, Sequencing, Lineage Placement, Demographic Analyses, and Cophylogenetic analysis of MAT loci

Taxon Sampling

Boletus edulis specimens were collected from a global distribution using both museum collections and targeted field sampling campaigns in Alaska, Utah, Germany, the United Kingdom, and Guatemala (Table S1). New specimens are deposited in the mycological collection of the Garrett Herbarium at the University of Utah. In addition, we included publically-available data from *Paxillus involutus* (Batsch) Fr., *Paxillus adelphus* J.-P. Chaumeton, Gryta, Jargeat P.-A. Moreau, *Hydnomerulius pinastri* (Fr.) Jarosch & Besl (Kohler et al. 2015) retrieved from the JGI Mycocosm Portal (Grigoriev et al. 2014), and the taxon sister to *B. edulis*, *Boletus reticuloceps* (M. Zang, M.S. Yuan M.Q. Gong) Q.B. Wang Y.J. Yao (NCBI accession PRJNA702108 (Dentinger et al. 2010; X. Wang et al. 2022)) (Q. Wang et al. 2005) as outgroups.

Whole Genome Sequencing

Genomic DNA was extracted from dried specimens using 10mg of hymenophore tissue. Tissue was first physically ground to a fine powder using 1 x 3.0 mm and 10-15 0.5 mm stanless steel beads in a 2.0 mL screwcap tube by shaking in a BeadBug^{TM} homogenizer set at 400 rmp for 30 seconds. Fully homogenized tissue was then used for DNA purification using the Monarch gDNA extraction kit (NEB biotechnologies). DNA extract quality was assessed for quality (260:230 nm and 260:280 nm wavelengths) using a NanoDrop 1000 spetrophotometer (Thermo Scientific) and DNA integrity was assessed using agarose gel electrophoresis. Samples were sequenced using a combination of paired-end sequencing on the Illumina MiSeq, HiSeq, and Novaseq sequencing platforms at the Genomics Core, University of Utah(Table S1). In addition, we generated additional long-read sequences using a combination of R9 and R10 Oxford Nanopore Flongle Flowcells for 21 specimens, and the Minion Flowcell for one specimen (BD747), to construct three reference-quality assemblies for each lineage (four for the AK lineage).

Genome Assembly, Phylogenetic Reconstruction, and Gene Prediction

For a visual representation of analysis pipeline and specimen filtering stages see supplemental flowchart (S1). Raw sequencing reads from 310 specimens were quality filtered and adapter-trimmed using fastP v0.20.1 (S. Chen et al. 2018) with default settings. Short-read genome assemblies were produced from quality-filtered reads using SPAdes v3.15.0 (Bankevich et al. 2012) with five k-mer values (k=77,85,99,111,127). To confirm the specimens as *B. edulis* and to classify them into lineages ,we used a summary coalescent phylogenomic approach. From each genome, we identified 1764 highly-conserved single copy orthologs using BUSCO (Waterhouse et al. 2018) with the "basidiomycota odb 10" dataset. Orthogroups that were present in less than 75% of specimens and specimens with less than 20% ortholog recovery were removed. Sequence alignment of retained orthologs was performed using MAFFT v7.397 (Katoh et al. 2017) with the L-INS-i algorithm, and maximum-likelihood gene-trees were inferred using IQ-TREE v2.0.3 (Minh et al. 2020) with automatic model selection in ModelFinder (Kalyaanamoorthy et al. 2017) and ultrafast bootstrapping (BS, Hoang et al. 2018) with 1000 replicates. A summary coalescent species tree was constructed from the resulting gene trees using ASTRAL-III (v5.7.5) after removing branches with very low support (BS<=10%) according to author recommendations (C. Zhang et al. 2018). The '*Boletus edulis*' group was defined as the monophyletic group sharing a most recent common ancestor with *B. reticuloceps*. Genomes that fell outside of the *B. edulis* and *B. reticuloceps* groups were removed from the dataset, and geographically monophyletic genetic clusters within *B. edulis* were identified as distinct lineages. After determination, we performed gene prediction of *B. edulis* specimens with at least marginal assembly quality as determined by a minimum BUSCO score of 50%. Gene prediction was performed utilizing the Funannotate (v1.8.9) (Palmer 2022) gene prediction wrapper script according to proscribed practices (Palmer 2022). To help guide gene prediction we used our previously published training set parameters constructed from *B. edulis*, (BD747, Tremble, Hoffman, et al. 2022), as well as previously generated mRNA sequences (Tremble, Hoffman, et al. 2022) mapped to each genome with hisat2 (V2.2.1) (Kim et al. 2019). To mitigate the potential for non-endogenous proteins in our pan-genomic analyses, we used the kmer based classification method kraken2 (V2.1.3, Wood et al. 2019) with a custom database (https://doi.org/10.7278/S50d-154b-fppf) to identify bacterial contigs within our genome assemblies. Only one sample possessed predicted proteins from non-endogenous bacterial contigs (UT90) and thus was removed from the pan-genomic analyses.

To enable high confidence comparative genomic analyses, we generated 21 additional reference-level hybrid assemblies. Three to four samples from each lineage, as defined from the phylogenomic analyses, were sequenced on R9 and R10 Flongle sequencing flowcells from Oxford Nanopore Technologies to supplement the Illumina short-reads with long-reads. Hybrid assemblies were generated with MaSuRCA (V4.0.9) (Zimin et al. 2013), and polished with POLCA three times (Zimin et al. 2013). Gene prediction was implemented with Funannotate as above. Kraken2 taxonomic classification was also conducted on each hybrid-assembly as above, but no non-endogenous scaffolds were identified.

Lastly, we sought to create a pseudo-chromosomal reference assembly for a single individual, which is necessary for any identification of large genomic re-arrangements. We generated additional long-read sequences for BD747 using a R10 MinION flowcell from Oxford Nanopore Technologies, and created multiple assemblies with different assemblers using Necat (Y. Chen et al. 2021), Nextdenovo (*NextDenovo* 2023), Smartdenovo (H. Liu et al. 2021), Flye (Kolmogorov et al. 2019), Canu (Koren et al. 2017), and Masurca (Zimin et al. 2013) with a minimum read-length of 1kb. Each individual assembly was polished three times with PILON (V1.24)(Walker et al. 2014), then a consensus assembly was generated using the RagTag pipeline (Alonge et al. 2022). Specifically, the assembly with the best BUSCO score, produced by FLYE, was scaffolded separately using each unique *de novo* assembly. A consensus assembly was then generated by identifying scaffold patterns produced by two or more assemblers with the merge function of RagTag. The final consensus assembly was again polished three times with PILON, followed by gene prediction with Funannotate as above.

Lineage Admixture and Demographic History Analysis

To genotype genome-wide single nucleotide polymorphisms (SNPs) within *B. edulis*, we first aligned adapter-trimmed and quality-filtered short reads to our newly created chromosomal reference with Bowtie2 (v2.2.6) (Langmead et al. 2012). SNPs were called with the GATK (V4.20) pipeline according to prescribed best practices. SNPs missing in more than 10% of samples, with a mean depth of coverage below 10, and with more than two alleles were removed from the dataset with VCFtools (v0.1.15) (Danecek et al. 2011). Lastly, samples with more than 30% missing data were removed from the dataset. To prevent the inclusion of putative clones in all analyses we used the –relatedness2 function of VCFtools to identify specimens with a high probability of clonality. If two or more sequences were identified as clonal, using the threshold of 0.354 according to author recommendation, we only retained the sequence with the best assembly BUSCO score, and other sequences were removed from the dataset for all downstream analyses.

Demographic histories were reconstructed using smc++ uv1.15.2 (Terhorst et al 2017), a method that takes advantage of both the site-frequency-spectrum (SFS) and linkage-disequilibrium (LD) under coalescent hidden Markov Chain Models. Regions interpreted by smc++ as ROHs larger than 10000bp were masked using the –missing-cutoff command. We used the mutation rate per generation of 2e-8 from another Agaricomycotina species (Baranova et al., 2015). To identify admixture within lineages we first used plink (V1.9) to remove any SNPs in high LD ($r^2 > 0.2$, 50Kbp window, 5bp step) from our dataset. Ancestry admixture analysis was performed with the snmf function of the R package LEA v3.2.0 (Frichot et al. 2014) on the LD-pruned dataset using sites with no missing values, with 10 runs for K=7. Lastly, we identified changes in predicted genome size and genome-wide heterozygosty using the kmer based estimation approach implemented in Genomescope with Jellyfish used for kmer identification.

2.1.1 Cophylogenetic analysis of maiting loci in *B. edulis*

To rule out sexual incompatibility as a driver of lineage segregation in *B. edulis* we performed a cophylogenetic analysis between the homeodomains from the fungal mating type loci (HD1 and HD2) and the species tree of *B. edulis*. Fungi exhibit non-associative mating, where sexual compatibility is determined by genetic diversity in a set of mating genes (MAT loci). Sexual reproduction in Fungi occurs randomly in a con-specific population when mating type loci are sufficiently different. If lineages are segregated through sexual incompatibility, then lineage relationships of the mating type loci should mirror that of the species tree.

Exonerate version 2.4.0 1 was used to extract homologs of the mating type loci HD1 and HD2. Homologs were extracted from assembled scaffolds using the –protein2genome and –bestn 2 flags to account for the dikaryotic nature of the samples. Query sequences of HD1 and HD2 from Rhizopogon salebrosus CDS (ON315867.1) were used as they were the most closely related publicly available sequences. Each set of homologs was aligned using the multiple sequence alignment program MAFFT version 7.475 2 with the parameters –maxiterate 1000 – localpair –reorder. Multiple sequence alignments were used as input for gene tree construction using IQTREE (v2.2.2.3) 3 with automatic model finder 4 and 1000 ultrafast bootstrap replicates optimized using the –bnni flag 5.

For cophylogenetic comparison, the species tree generated for *B. edulis* had tree branches pruned using the drop.tip function of APE (5.7-1) to match the alignment input for each gene tree. Cophylogenetic comparison between the mating type loci gene trees and species tree was performed using the “cophylo” function of the R package Phytools (1.5-1)7 with the flag methods=c("pre","post") and “roatate.multi=T” and to automatically rotate and detangle branches.

2.2 Identifying Macrogenomic variation in *B. edulis*

Characterizing the Pangenome of *B. edulis*

Segregation among lineages of *B. edulis* may be maintained by large shifts in gene repertoire, rather than by small changes in the abundance or divergence of adaptive loci. To investigate whether lineages in close geographic space possess distinct gene repertoires, we constructed the global pan-genome of *B. edulis* and characterized the distribution of accessory proteins within and between lineages. First, we conducted gene prediction on all genomes with at least marginal assembly quality (minimum BUSCO score $> 50\%$), 155 genomes in total, using Funannotate as descibed above. To assess investigate the impact of assembly quality on pan-genome prediction we used predicted protein sequences from all 155 *B. edulis* specimens with at least marginal assembly quality (BUSCO scores greater than 50%) to predict orthologous genes with Orthofinder using the "–diamond-ultra-sens" search algorithm (Emms et al. 2019). All genes not placed in orthogroups (excluding singleton genes placed in their own orthogroup) were removed from subsequent analyses. Genes were classified as "core" if they were present in 95% of genomes, "accessory" if they were present in at least

two gnomes but in less than 95% of them, and "singletons" if they were present in only one genome. After determining that minimum assembly quality had a significant impact on the abundance of predicted core proteins, we then filtered our dataset to include only 92 high quality assemblies, (>=85% BUSCO recovery), and reclassified orthogroups as core, accessory, and singleton. To test for the clustering of accessory gene profiles by lineage, we calculated Bray-Curtis distance between all samples and used PERMANOVA to determine significance (adonis command of R package Vegan (v2.6.4)). To visualize lineage clustering based on accessory gene profile, we implemented a principal coordinate analysis of our Bray-Curtis distance matrix using the R package "ecodist".

Identifying Large Structural Variation and Changes in Genomic Synteny

To identify large-scale structural variation across the *B. edulis* lineages we used two independent bioinformatic approaches to predict large structural variants, 1) a highly accurate approach using whole-genome mapping of our reference quality genomes to first identify potentially adaptive variants, and 2) a more sensitive long-read nanopore mapping approach to verify putative adaptive variants. These approaches combine the accuracy needed to identify variants confidently and prevent false-positive identifications (1), while retaining the sensitivity needed to identify low-frequency variants (2) and prevent false-negative identifications. First, we performed whole genome alignments of our 21 hybrid-reference genomes against the 10 largest chromosomes of our pseudo-chromosomal reference genome using minimap2 (V2.23) with the -x asm20 command. We chose to target only the largest 10 scaffolds of our pseudo-chromosomal reference, as the scaffold size of additional scaffolds decreases dramatically and many of these are likely not pseudo-chromosomes. Inversions and translocations greater than 1Kbp in length were predicted from alignments using the accurate MUM&Co prediction pipeline (O'Donnell et al. 2020). Variants found in at least two genomes of the same lineage were classified as potentially adaptive. Second, we used the sensitive SNIFFLES2 pipeline to predict large structural variants using long-read nanopore sequences mapped to the 10 largest pseudo-chromosomes of our reference with minimap2 (-x ont command). We used these additional predictions to verify the potentially adaptive variants identified with the first approach and only retained variants with evidence of presence in all three/four genomes of a lineage.

To determine if coding regions found within structural variants are hot-spots of genomic divergence, we calculated the population fixation index F(st) on a pairwise basis between lineages using PIXY (v1.2.7.beta1). We reduced our SNP dataset to only include predicted exons (CDS sequences) and estimated F(st) across only exons within predicted structural variants, and across exons genome wide.

To estimate changes in genomic structure between lineages in a pairwise manner, we identified the proportion of genes within any two reference genomes that have syntenic flanking regions. First, we identified all orthologous gene groups across our 22 reference genomes with orthofinder. We then identified all single-copy orthologous groups, and recorded the

orthogroup ID of the gene immediately upstream and downstream of each single-copy gene in all genomes, excluding single-copy genes at the beginning or end of scaffolds. A single-copy gene was considered "syntenic" between two genomes if the orthogroup IDs of the upstream and downstream genes matched regardless of order. Finally, we identified the proportion of single-copy genes shared between two genomes in all pairwise combinations (21 comparisons per genome) that were considered syntenic.

2.3 Identifying microgenomic changes in *B. edulis*

Functional Protein Analysis

To identify whether lineages of *B. edulis* possess distinct repertoires of ecologically relevant gene families, we characterized the saprotrophic enzyme profiles of the *22* high-quality reference genomes. Carbohydrate activate enzymes ("CAZymes"), which are implicated in primary metabolism from decomposition, were identified from predicted proteomes using the run_dbCan4 program (H. Zhang et al. 2018) and HAMMER (Finn et al. 2011), DIAMOND (Buchfink et al. 2015), and ECAMI tools (Xu et al. 2020). Putative CAZymes were characterized as genes with significant hits using at least two of the three methods. Peptidases (quick description) were identified by PHMMER (V3.3.2) searches of predicted protein sequences against the MEROPS peptidase database (Rawlings et al. 2018). Significant hits were classified as Evalue $< 1e^{-10}$, and only the significant hit with the lowest Evalue was chosen for each predicted protein.

Effector protein analysis

To identify whether lineages of *B. edulis* differ in their likely capacity to colonize varied hosts, we characterized the abundance and diversity of effector proteins. Effector proteins are well-known to be the primary group of secreted proteins that facilitates host colonization in pathogenic fungi (Uhse et al. 2018), and limited work in EMF has also shown that secreted effectors are necessary for successfully establishing mycorrhizae (Jonathan M. Plett et al. 2011). Secreted proteins were predicted from the proteomes of our 22 *B. edulis* reference genomes, using the SignalP6 program in slow-sequential mode (Teufel et al. 2022). Effector proteins were characterized from predicted secreted proteins using effectorP 3.0 (Sperschneider et al. 2022). EffectorP identifies secreted proteins that possess characteristics commonly found in known fungal effectors, such as being low molecular weight, having a high abundance of cysteine residues, and C-terminal regions enriched in lysine. To identify accessory and core effectors across all genomes we used orthofinder (Emms et al. 2019) to identify orthologous groups, and Principal component analysis (R package APE v5.7) was used to cluster genomes based on their effector protein profiles.

Abundance and Diversity of Transposable Elements

The genomes of ectomycorrhizal taxa have been consistently shown to exhibit significant expansions in repetitive DNA content (Min et al. 2020; Wu et al. 2022), and it is thought that

transposable elements (TEs) may play a role in the rapid diversification of some adaptive gene families such as effector proteins. To characterize TE content, we used the EDTA (Ou et al. 2019) wrapper script to predict and annotate TEs in our set of 22 reference-genomes to determine if each of the *B. edulis* lineages have experienced unique expansions or contractions in TE content. Repeat-rich regions are commonly the most difficult genomic regions to assemble, and thus the prediction of TEs is likely to be greatly impacted by even small differences in assembly quality. Therefore, to account for differences in the assembies, we used the number of scaffolds to scale the abundance of predicted TEs as follows;

$$T_b = T_a(log(S)) \tag{1}$$

where S=number of assembled scaffolds, Ta= predicted TE abundance, and Tb=scaled TE abundance. To determine whether TEs may be clustered around potentially adaptive genes, we calculated the mean number of TEs present in a 50kbp window centered on all predicted effector proteins, secreted proteins, CAZymes, peptidases, and all other proteins across the gneome that did not fall within one of these categories. To test whether the distribution of TE density for each class of enzyme was significantly different than the genome-wide average, we used the two-sided Kolmogorov-Smirnov test in R (ks.test).

3 Results

3.1 Genome Sequencing, Lineage Placement, Demographic Analyses, and Cophylogenetic analysis of MAT loci

Genome sequencing and lineage placement

From 310 initial specimens, genome sequences from 262 (85%) produced assemblies with more than 20% BUSCO recovery (Table S2). Summary coalescent analysis of 1733 gene-trees with ASTRAL-III, revealed seven monophyletic *B. edulis* clades with 100% statistical node support that share a most recent common ancestor with the Asian endemic *B. reticuloceps*. Six of these lineages directly correspond to the six lineages previously identified (Tremble, Hoffman, et al. 2022), and hereafter will be referred to as AK (Alaska/Siberia, BC (British Columbia), CO (Colorado), EC (Eastcoast North America), EU (Europe), and WC (Westcoast North America). The seventh lineage is a novel group that is exclusively comprised of tropical *B. edulis* specimens from Guatemala and will henceforth be referred to as GU (Guatemala). GenomeScope analysis of genome size and genome heterozygosity identified significant differences among lineages (Fig. S8C,D). Average genome size across all of *B. edulis* was 41.2 Mbp, and average genome heterozygosity was 1.22%. The GU lineage was the least heterozygous (0.623%) and had the largest predicted genome size (47.0 Mbp) ($P < 0.05$, ANOVA, Fig. S8C).

After lineage placement, we sought to produce three reference genomes per lineage

(four for the widely distributed AK lineage) for the accurate assessment of gene-family abundance and evolution. Hybrid-sequencing produced 21 highly-contiguous reference genomes (mean N50 = 225 Kbp, mean number of scaffold = 712, Table S3). In addition, we re-sequenced our original reference genome from sample BD747 of the WC lineage and produced a pseudo-chromosomal reference assembly. MinION sequencing produced, 9.85 Gbp (234X coverage) of long-reads, with a library N50 of 5.49Kb. The final consensus assembly is 44.6 Mbp long, contains 65 scaffolds, and a N50 of 2.16 Mbp..

Estimating admixture among lineages of *B. edulis*

Admixture analysis reveals low levels of admixture among most of the lineages, despite parapatric and near sympatric distributions (Fig. 1 A,B), confirming earlier results (Tremble, Hoffman, et al. 2022). However, the amount of admixture varied greatly among the lineages and when present was often unidirectional. For example, the AK lineage contains the greatest abundance of admixed individuals, and the source of this admixture is dependent on geography, with Alskan individuals of the AK lineage showing evidence of admixture with the EC lineage, and western Russian individuals of the AK lineage showing evidence of admixture with the EU lineage. However, the EC and EU lineages showed substantially lower levels of admixture with the AK lineage, implying that gene flow is more or less unidirectional into AK.

Estimating historic changes in effective population size within lineages of *B. edulis*

Modeling of historic changes in the effective population size (N_e) show that demographic histories vary appreciably among the lineages (Fig. 1 C). With the exception of the GU lineage, all of the groups exhibit a substantial reduction in N_e around 30,000 generations ago. The GU lineage exhibits a more ancient drop in N_e around 50,000 generations ago. Overall, the BC, EU, and WC lineages show similar trends in N_e, with slightly greater demographic stability in the EU lineage. Interestingly, only the GU lineage exhibits a recent increase in N_e, whereas all of the other lineages exhibit decreases in N_e in recent time.

Cophylogenetic analysis of MAT loci in *B. edulis*

To investigate if lineages of *B. edulis* are isolated by sexual incompatability, we compared phylogenetic relationships based on the the species tree and mating type loci. Cophylogenetic comparison between HD1 and HD2 mating type homologs and the species tree of *B. edulis* showed incongruence across all lineages for both genes (Fig. S3). The lack of population structure in the mating type loci is consistent with expectations for lineages capable of interbreeding.

3.2 Identifying structural variation across *B. edulis*

Characterizing the Pangenome of *B. edulis*

To determine whether genomic divergence is maintained by large shifts in total gene repertoires we characterized the global pangenome of *B. edulis*. First, we characterized the impact of assembly quality on core protein prediction by including all genomes with at least marginal assembly quality (minimum BUSCO recovery = 50%) for a total of 155 genomes. We found that assembly quality had a large impact on the number of core proteins identified from a minimum BUSCO score of 50% to a minimum BUSCO score of 80% (Fig. 2A). At the 80% minimum BUSCO score threshold, a stable plateau of core protein abundance was reached, and less than 1% variance in the abundance of core proteins was observed when quality thresholds were increased from 80% (7,091) to 85% (7,142) to 90% (7,194) BUSCO recovery. Subsequently, we chose the 85% BUSCO score as our minimum assembly quality threshold for inclusion in the final characterization of the *B. edulis* pan-genome, resulting in a final total of 92 genomes. This threshold represents a balance between including a sufficient number of genomes to capture the total diversity of gene families while removing genomes prone to false negatives due to assembly fragmentation. Using this threshold we identified 21,142 putative gene families; 33.8% (7,142) of these gene families were core proteins, 52% (10,977) were accessory proteins, and 14.3% (3,023) were singleton proteins (Fig. 2B). The abundance of core proteins found within a single lineage was relatively uniform across the group (min. core proteins, GU = 6,962, max. core proteins, CO = 7,450, mean = 7,137, Fig. 2C). However, the abundance of accessory proteins varied considerably (min. acc. proteins, CO = 2,111, max. acc. proteins EC = 4,467, mean = 3,194 Fig. 2C), where the lineage with the most accessory proteins (EC) contained over twice the abundance of accessory proteins as the lineage with the fewest proteins (CO).

Given the strong geographic structure and high genomic divergence of *B. edulis*, we hypothesized that many accessory proteins would be lineage-specific. However, only eight (0.1%) of the predicted accessory proteins were found to be exclusive to and abundant within (found within 95% of genomes) any one lineage and potentially adaptive (AK = 2 gene, GU = 5 genes, WC = 1 gene). Only one of these lineage-specific accessory proteins had a predicted function: OG0013714, which is found in the GU lineage and has predicted homology with pyroxidal kinase (IPR004625), which is the enzyme responsible for phosphorylation and activation of vitamin B6, a well studied antioxidant in fungi (Samsatly et al. 2020).

To investigate the geographic distribution of accessory proteins within *B. edulis*, and to identify whether accessory proteins cluster by geography, lineage, or both, we conducted a principal component analysis of the genomes based on their accessory protein profiles. We identified four distinct pan-genome clusters that do not correspond with phylogeny (Fig. 2D). For example, the AK and EC groups are phylogenetic sister-lineages, yet the AK lineage contains a protein profile more similar to the EU group than to the EC group. In addition, the BC and WC lineages form a distinct pan-genome cluster, despite the WC lineage sharing a more recent common ancestor with the CO and GU lineages. This finding suggests that

evolutionary history or admixture alone cannot predict whole genome proteome structures.

Identifying Large Structural Variation and changes in Genomic Synteny

To identify whether genomic structural variation is necessary to maintain lineage segregation in *B. edulis*, we predicted the abundance and composition of large structural variants within the 22 reference-quality genomes. In total, we identified 252 predicted inversions (24.8 Kbp average length) and 1056 translocations (38.2 Kbp average length) using whole genome mapping, of which 103 (41% of total) and 125 (12% of total) were found in at least two individuals respectively (Fig. 3A, Table S9). After controlling for assembly size, we found a similar abundance of inversions and translocations (0.3 INV/Mb of genome, 0.15 TRA/Mb of genome) in *B. edulis* to that of yeast (*Saccharomyces cerevisiae* 0.3 INV/Mb of genome, 0.18 TRA/Mb of genome (Bendixsen et al. 2021)), and cacao (*Theobroma cacao* 0.24 INV/Mbp, (Morin et al. 2017)). Surprisingly few inversions or translocations were shared between individuals of the same lineage and were classified as potentially adaptive: 62 (25% of total), and 47 (4%) respectively. After verifying these potentially adaptive variants with additional evidence generated from long-read nanopore mapping we identified only 9 inversions and 18 translocations that were found within all three (or four for the AK lineage) genomes of a lineage. Some lineages, particularly CO, EC, and GU, possessed multiple potentially adaptive structural variants (Table S4), while others, AK, BC, and WC, exhibited few to none. In addition to being more abundant, translocations tended to be substantially larger than inversions (tranlocation mean size = 58 kb, inversion mean size = 16.5 kb), although this could be due to the greater difficulty in identifying small translocations.

To determine whether protein coding-regions located within potentially adaptive structural variants are diverging at a faster rate than the rest of the genome, and potentially contributing to lineage segregation, we calculated pairwise F(st) values across predicted exons genome wide and compared this to F(st) values of exons within structural variants. We found that both inversions and translocations can greatly impact the rate of divergence, but that this impact varies appreciably among different structural variants and lineage comparisons. In some comparisons, protein-coding regions within structural variants exhibited far more divergence than the rest of the genome, but only when compared with a sister lineage. For example, we identified an inversion in the CO lineage that is potentially adaptive (Fig. 3B). F(st) values of the coding regions in this inversion are much greater than the rest of the genome only when comparing the CO lineage with its sister lineage, the GU lineage. In all other comparisons, the coding regions of this inversion are less divergent than the genome background. In contrast, some structural variants exhibit no obvious differences in divergence rates compared to the rest of the genome for all comparisons (Fig. 3C).

To directly compare the degree of synteny between any two reference-level assemblies we calculated the proportion of shared genes with syntenic flanking regions on a pairwise basis. On average, 3,092 genes (min = 2,146, max = 3,690) were used to test for synteny between any two genomes, with 55.5% (min = 45.1%, max = 60.7%) of these being identified

as syntenic on average. Similar levels of synteny were found between individuals of different lineages as compared to individuals of the same lineage. The mean proportion of syntenic genes did not differ significantly between within-lineage comparisons and between-lineage comparisons (Fig. S7, mean within = 0.5585, mean between = 0.5544, P = 0.345 ANOVA). Within a lineage, the average proportion of syntenic genes was found to vary substantially (Fig. S6), but not within all of the lineages. For example, the EU specimens all possess similar levels of synteny, while the BC and WC lineages each contain individuals with distinctly lower synteny. Despite this lack of conservation, significant differences in the average proportion of syntenic genes were found among lineages; the EC lineage exhibits the lowest degree of synteny (P « 0.001, ANOVA), while the EU possess the greatest degree of synteny among all of the lineages.

3.3 Identifying microgenomic changes in *B. edulis*

Functional Protein Analysis

We quantified the abundance and diversity of genes encoding CAZymes and peptidases within the set of 22 reference-quality genomes to investigate if the lineages of *B. edulis* differ in their saprotrophic enzyme repertoires. We identified 100 CAZyme families and 83 peptidase families across all lineages of *B. edulis* (Table S6). Of these, 90% of CAZyme and 92% (76 of 83) of peptidase families were found in all of the specimens (Fig. 4A). No enzyme families were lineage specific. However, we did find substantial variation in the abundance of saprotrophic enzymes within and among individual lineages (Fig. 4A). In particular, certain enzyme families showed significant losses and expansions that were unique to a specific lineage. For example, the WC lineage possessed the greatest abundance of metallopeptidases ($P < 0.001$ ANOVA, Fig. 4D), and the cellulolytic CMB1 enzyme family was entirely lost from the GU lineage. Across all enzyme families, the GU lineage showed consistent contractions in most saprotrophic enzyme families and possessed the fewest enzymes in total ($P < 0.001$, ANOVA, Fig. 4A, Table S6). PCoA clustering of saprotrophic gene repertoire further highlighted the unique saprotrophic composition of the GU lineage (Fig. 4 C) which is highly segregated from all other lineages. Interesting, the WC and BC lineages exhibit no PCoA overlap suggesting that these lineages possess distinct saprotrophic repertoires despite persisting in parapatry. PCoA clustering also indicates that the presence of admixture is not a good predictor of saprotrophic repertoire similarity. The AK lineage possessed a substantial degree of admixture with both the EC and EU lineages, yet only exhibited saprotrophic PCoA overlap with the EU lineage.

Effector Proteins

We quantified the diversity and abundance of all effector proteins present in the seven lineages. In total, we identified 289 orthologous effector groups, of which 58 were core to. Surprisingly, we found a relatively high degree of effector conservation among the lineages, with 67%-85% of effectors being shared between any two lineages. Moreover, we found little

evidence for lineage-specific divergence, as none of the effector proteins were found to be unique to and abundant within any one lineage. The overall abundance of effector proteins within *B. edulis* followed a bimodal distribution (Fig S9), with most of the proteins being present in either two genomes or all of the genomes. Despite this conservation, we identified significant differences among the lineages in the total abundance of predicted effectors, and particularly the abundance of predicted cytoplasmic effectors (Fig. S8A). The EU lineage possessed the greatest average abundance of both cytoplasmic (112) and total effector proteins (311) ($P < 0.0001$, ANOVA) (Table S7). Principle component clustering of effector protein profiles revealed two main clusters comprised of multiple lineages and two clusters of individual lineages (Fig. 4B). The BC, WC, and EC lineages comprised one distinct cluster, while GU and CO comprised the second distinct cluster. The AK and EU lineages did not share overlap with any other lineage.

Abundance and diversity of Transposable elements

On average, repetitive elements comprised 8.2% of the genome (min = 3.36%, max = 14.67%), and were primarily comprised of long-terminal-repeat (LTR) Gypsy (3.02%), LTR Copia (2.3%), and terminal-inverted-repeat (TIR) TE families (1.41%). After controlling for genome assembly fragmentation, we found highly significant differences in the abundance of repetitive content per lineage (Fig. S8E). The GU lineage exhibits the largest expansion of TEs, while the BC lineage carries the least TEs. Across all of the genomes, we found that TEs were significantly less abundant in 50kb windows around secreted proteins than across the rest of the genome ($P < 0.001$, Kolmogorov-Smirnov test, Fig. S10), and showed a marginally significant tendency to be more abundant in 50kb windows around effector proteins ($P=0.08$, Kolmogorov-Smirnov).

4 Discussion

Here, we sought to identify whether lineage segregation within a globally distributed, non-model eukaryote with high levels of gene flow, is maintained by large genomic structural variation, reinforcing local adaptation through hybrid incompatibility, or by smaller changes in the abundance of specific ecologically relevant gene families subject to local selection. To this end, we characterized the global pan-genome of *B. edulis*, inferred the abundance, diversity, and divergence of large inversions and translocations, and quantified the saprotrophic and effector gene repertoires of its phylogenetic lineages. We found little evidence that would suggest structural variation plays a significant role in maintaining lineage segregation. After characterizing the exceptionally diverse pan-genome of *B. edulis*, we found that lineages that exhibit little to no admixture despite a parapatric distribution can exhibit substantial pan-genome overlap. In contrast, lineages with strong signatures of gene flow with substantial admixture have little pan-genome overlap and form distinct genetic clusters. Furthermore, we found that translocations and inversions are abundant within *B. edulis* but that these variants are rarely shared between any two genomes. When they are shared, this tends

to occur as often between genomes of separate lineages as between genomes of the same lineage. By implication, gene flow within and between lineages can persist despite much structural variation, and at the same time, that gene flow can be interrupted in the absence of substantial structural variation.

In contrast, we found that lineages of *B. edulis* possess distinct saprotrophic and effector protein repertoires, even when they are in close geographic proximity. The lineage from at the lowest latitudes of its natural range (GU) is the most dramatic example of this, displaying significant contractions and losses of most saprotrophic gene families, and is highly segregated from the rest of *B. edulis* in the PCoA. Moreover, even lineages with less dramatic changes in saprotrophic gene repertoire exhibit consistent differences. For example, the BC and WC lineages exhibit the greatest amount of geographic overlap, yet they do not overlap in their saprotrophic repertoires. This is likely due to the large expansion of metallopeptidases in the WC lineage. This suggests that the evolution of saprotrophic genes in *B. edulis* is dynamic, and that these gene families may be actively involved in local adaptation. Lineages of *B. edulis* also exhibit distinct effector protein repertoires. Importantly, some lineages that overlap in their saprotrophic repertoires, do not show parallel overlaps in their effector repertoires. For example, the AK and EU lineages, which exhibit substantial levels of admixture, greatly overlap in their saprotrophic repertoires, but possess distinct effector protein profiles.

4.1 The *B. edulis* pan-genome reveals unprecedented genomic diversity

The pan-genome of *B. edulis* is one of the most diverse pan-genomes currently known and contains the greatest diversity of accessory proteins and the lowest abundance of core proteins identified in *Fungi*. The core genome of a species can be comprised of only 33% - 87% of all identifiable genes in plants (Bayer et al. 2020), and 55% - 90% in *Fungi* (Badet et al. 2020; Kaushik et al. 2022; Lofgren, Ross, et al. 2022; McCarthy et al. 2019). Here, we find that only 33.8% (7,142) of predicted proteins in *B. edulis* are considered core. This low abundance of core pan-genome proteins could also be due to our having treated *B. edulis* as a single species. Due to its wide geographical, ecological, and morphological diversity, *B. edulis* has been variously considered as a single widespread species or multiple species using inconsistent lines of evidence and piecemeal geographic sampling that suffers from severe ascertainment bias (Arora 2008; Arora and Frank 2014; Dentinger et al. 2010; Smith 2005; Tremble, Suz, et al. 2019, 2020). Tremble et al. (2022) showed that high rates of gene flow are detectable across all lineages of *B. edulis* despite strong geographic structure with limited admixture, reinforcing the argument that an inclusive classification more accurately reflects the underlying biology. Thus, while the comparatively low percentage of core proteins identified in *B. edulis* could be due to the inclusion of a more divergent set of genomes than that of a well regarded "species", such as *Aspergillus fumigatus* (Lofgren, Ross, et al. 2022), even at the scale of an individual lineage, *B. edulis* contains a lower percentage of core proteins (46.5% core in the AK lineage) than has been previously identified in *Fungi*.

Moreover, if *B. edulis* was actually comprised of seven reproductively isolated, independently evolving species, we would expect to find a much greater abundance of lineage-specific core proteins than only the eight we identified (Hyun et al. 2022; Khan et al. 2020). Therefore, we hypothesize that the low abundance of core proteins found within *B. edulis* is the product (or cause) of its exceptional ecological amplitude.

Lineage boundaries, gene flow, and admixture can be maintained despite large proteome-wide divergence. We found significant clustering based on accessory protein profiles, with four clear clusters correlating with geography and not phylogeny. If genome-wide restructuring of the proteome is necessary for long-term evolutionary divergence we would expect lineages in close proximity to exhibit distinct differences in pan-genome clustering. However, the two lineages in near sympatry, BC and WC, form a strong pan-genome cluster, and some individuals are more clustered with members of the opposite lineage than their own. Despite this pan-genomic similarity we still find little admixture between the BC and WC lineages. We also found evidence of substantial admixture from the EC lineage into AK, which aligns with our previous findings that the AK and EC lineages exhibit the strongest ongoing signals of gene flow, and the lowest levels of nucleotide divergence (Tremble, Hoffman, et al. 2022). However, pan-genome clustering indicates far greater pan-genome similarity between the AK and EU lineages than the AK and EC lineages. Thus, *B. edulis* carries exceptional pan-genomic diversity, with lineage boundaries being maintained in the absence of proteome-wide divergence.

4.2 Structural variation in *B. edulis* is abundant, yet rarely shared

Genomic architecture plays an indisputably important role in evolution, with large structural variants having been widely implicated in the process of divergence (Bendixsen et al. 2021; Hämälä et al. 2021; L. Zhang et al. 2021). However, it is unclear to what extent structural variation can innocuously persist within a group, and whether structural variants are necessary for long-term maintenance of lineage boundaries. For example, a recent genomic survey of the order *Boletales* found that loss of synteny accrues with phylogenetic distance, but not uniformly across the genome, suggesting that many structural variants only accumulate in regions of low constraint (Wu et al. 2022). In contrast, many ascomycete fungi often lack conserved genomic synteny when comparing any two genomes (termed mesosynteny within *Dothidiomycetes*), and can possess multiple accessory chromosomes (Goodwin et al. 2011), indicating that structural variation may not be an insurmountable post-zygotic barrier. Here we found that ample structural variation persists within *B. edulis*, but that most of the variants result from isolated events occurring within single individuals. When structural variants were found in more than one individual, they tended to be present in individuals of different lineages as often as they were found within individuals of the same lineage. Moreover, only three out of seven lineages exhibited a large structural variant found within all individuals of the lineage. These potentially adaptive variants constituted only 1% of all predicted variants. When checking for more localized changes in gene-order synteny we found that the degree of synteny was similar for any two genomes of differing lineages as it was for two genomes

of the same lineage. To the best of our knowledge, the only other species group within the *Basidiomycota* that has been surveyed for genomic synteny and structural variation within the context of species boundaries is the *Cryptococcus amylolentus* and *Tsuchiyaeawingfieldii* species complex. Interestingly, this group exhibits nearly opposite patterns of synteny to those of *B. edulis*. Passer et al. (2019) revealed substantial conserved genomic synteny between isolates. However the few structural variants that were identified, in particular a single large reciprocal translocation, may be contributing to postzygotic barriers through hybrid-incompatibility (Passer et al. 2019). As only four parental strains were available at the time of study, it is possible that structural variation is actually more common within the *Cryptococcus amylolentus* complex than identified in Passer et al. (2019). Regardless, it is clear that the role of structural variation in contributing to species boundaries is vastly different in the *Cryptococcus amylolentus* complex than in *B. edulis*. Therefore, without additional surveys of synteny across the *Basidiomycota*, it remains unclear whether the abundance of structural variation within *B. edulis* is common or unique within the phylum.

However, these results do not imply that structural variation is altogether not contributing to divergence in *B. edulis*. We found that some protein coding regions within potentially adaptive inversions and translocations can exhibit greater levels of divergence than the rest of the genome (Fig. 3B). Yet, these coding regions only exhibited high divergence when compared to a single specific lineage, and not across all lineage comparisons. Furthermore, some variants were no more or less divergent than the rest of the genome (Fig. 3C). This suggests that structural variants contribute to divergence in *B. edulis* by sheltering locally adaptive loci from recombination, and not by initiating divergence by themselves. This is consistent with a growing body of evidence concerning the role of structural variation, particularly inversions, in local adaptation (Tigano et al. 2016; Wellenreuther et al. 2019) It is important to note, in *B. edulis* it appears that structural variants can facilitate divergence within potentially locally adaptive loci, but that they are not explicitly necessary for local adaptation. All lineages within *B. edulis* exhibit strong signatures of local adaptation (Tremble, Hoffman, et al. 2022), and here we provide even further evidence to support this hypothesis. However, we found no evidence of widespread potentially adaptive structural variants in either the EU, BC, or WC lineages, suggesting that local adaptation can proceed in the absence of local sheltering from recombination. Altogether, we found little evidence to suggest that large structural variants play a significant role in maintaining lineage fidelity within *B. edulis*. Much variation can persist within a single lineage, indicating this variation contributes little to reproductive barriers. Lastly, when structural variants do contribute to divergence, they facilitate the divergence of locally adaptive loci; i.e. loci that may have become divergent regardless of the genomic context.

4.3 Adaptation to diverse ecological pressures may be driving divergence in *B. edulis*

To investigate if evolution in saprotrophic and effector protein gene families occurs over short evolutionary time scales, and if differences in gene family abundance or diversity may con-

tribute to the maintenance of lineages in *B. edulis*, we generated 22 high-quality reference genomes and conducted a global comparative genomic survey. We identified three distinct patterns of ecological divergence that may be underly lineage segregation in *B. edulis*. First, we found that the GU lineage has experienced significant contractions in decomposition related gene families and possesses a saprotrophic gene repertoire distinct from the other lineages. Moreover, it appears that this loss represents a consequential niche shift in the GU lineage, which uniquely exhibits demographic bottleneck, significantly reduced heterozygosity, and expansions in genome size and TE content. It has long been proposed that tropical soils represent a unique habitat for EMF compared to temperate soils (Tedersoo, Bahram, et al. 2012; Tedersoo and Nara 2010). This is because high decomposition rates lead to rapid nutrient cycling and low levels of soil organic matter in tropical soils (Jenny et al. 1949; Ladd et al. 1985). Moreover, a greater abundance of sunlight hours and precipitation in tropical latitudes increases the photosynthetic capacity of EMF hosts, increasing the abundance of readily available carbon that can be allocated to EMF (Kuikka et al. 2003; Kummel et al. 2011). These factors may reduce the need for EMF to enzymatically liberate nitrogen and other nutrients from soil organic matter, or the need to acquire carbon from non-host sources during periods of drought or low sunlight, subsequently yielding a contraction in saprotrophic machinery. A shift in the preferred host might also generate a similar pattern, as hosts differ in their photosynthetic capacity. However, this is unlikely the case in *B. edulis* as the GU and CO lineages possess very similar effector protein profiles despite associating with different hosts, yet exhibit vastly different saprotrophic repertoires.

The second pattern we identified is that the EU lineage exhibits a unique expansion of effector proteins, potentially a signature of a true host generalist. While differences in host association in *B. edulis* are unclear, it is thought that European *B. edulis* are broad generalists, and can equally colonize both angiosperms and gymnosperms. In fact, mycorrhiza between European *B. edulis* and diverse hosts have been formed in laboratory settings (Águeda et al. 2008), supporting this hypothesis. In contrast, it is thought that *B. edulis* from North America are primarily restricted to gymnosperms, with the vast majoirty of observatios recorded in Pinaceae-dominated forests and with confirmed angiosperm associations being relatively rare (MyCoPortal. 2023. http://www.mycoportal.org/portal/index.php. Accessed on August 28). This is supported by the absence of any confirmed associations with *Fagaceae* despite their ecological dominance throughout much of eastern North America. Therefore, the expansion in effector proteins in the EU lineage, particularly the expansion in cytoplasmic effector proteins, may provide further evidence that European *B. edulis* have an increased capacity to colonize a broad array of hosts. However, it is unknown how the composition of effector proteins within a genome impacts the capacity to colonize particular hosts. For example, a recent study of effector protein abundance and diversity in the ectomycorrhizal genus *Suillus* found little difference in the total abundance of effector proteins between host specialists and host generalists (Lofgren, N. H. Nguyen, et al. 2021), although its host range is far more restricted than in *B. edulis*. Surprisingly, we identified no abundant lineage specific effector proteins in *B. edulis*, suggesting that if differences in host association exist, they likely cannot be explained by differences in capacity. However, colonization preference might

be also driven by other molecular modulators, such as microRNA's (Wong-Bajracharya et al. 2022), which were not characterized in this study. Interestingly, the neighbor lineages of AK and EU, which exhibit relatively high levels of admixture (predominately in the direction of EU to AK), possess large overlap in saprotrophic repertoire, yet possess distinct effector protein profiles. In contrast, differences in saprotrophic capacity, but not host preference, may maintain lineage segregation between CO and GU. Thus, divergence between AK and EU appears to be driven by host preference, whereas divergence between CO and GU appears to be driven by saprotrophic capacity. This suggests that lineage segregation in *B. edulis* can be maintained by multiple, distinct context-dependent ecological pressures.

Finally, we found that the WC lineage exhibits a significant expansion of metallopeptidases, consistent with expectations from adaptation to distinct soil conditions. Extracellular peptidases are among the most important enzymes used by EMF for the degradation of soil-organic matter (Bending et al. 1995; Shah et al. 2013), and are primarily comprised of serine- and metallo-peptidases in *basidiomycete* fungi (Semenova et al. 2017). Recent work has also shown that peptidase activity varies among soils collected from multiple sites (T. Nguyen et al. 2019), and that separate peptidase families are preferentially active in distinct pH ranges (Semenova et al. 2017). The WC and BC lineages are parapatric in distribution (nearly sympatric in Washington state), and exhibit a high degree of overlap in pan-genome clustering and effector protein profiles, while they are not particularly distinct in their overall saprotrophic repertoires. However, the WC lineage exhibits a significant expansion in metallopeptidases compared to the rest of *B. edulis*. This expansion may suggest that the WC lineage has adapted to unique soil conditions compared to the BC lineage, and that small, relatively cryptic changes in ecological preference may be sufficient to maintain lineage fidelity.

5 Conclusion

The mechanisms that maintain stable lineage boundaries through time are poorly understood. It is unclear whether divergence generated by local adaptation needs additional reinforcing changes, such as large chromosomal rearrangements that generate hybrid-dysfunction, to maintain segregation. In *B. edulis*, we found that lineages in close geographic proximity (near sympatry) exhibit significant pan-genome overlap, few to no lineage-specific structural variants, and no significant loss of synteny. Despite this genomic conservation, these lineages exhibit little admixture, indicating stable boundaries. Nevertheless, we found significant differences in the abundance and diversity of saprotrophic enzymes, and in the abundance of effector proteins between lineages. Based on these results, we hypothesize that the EU lineage of *B. edulis* has experienced a host expansion, the GU lineage is more dependent on host derived carbon and/or is utilizing more labile soil-organic material, and the WC lineage has adapted to distinct soil conditions. Altogether, our results suggest that structural variation, in of itself, is insufficient to maintain lineage segregation in *B. edulis*, and that diverse forms of local adaptation and subtle differences in ecological preference can maintain stable

lineage boundaries across the 2.6 million years of divergence in *B. edulis*.

6 Acknowledgements

The authors would like to acknowledge the many institutions and individuals who provided specimens for use in this work (See Table S1 for a full list of collaborating holding institutions). We would like also like to thank Tiffany Do and Sariah VanderVeur for their effort in extracting genomes for this work.

7 Data Availability

All short-read genome sequences, assemblies, and predicted protein sequences as well as all 22 reference genome assemblies and predicted protein sequences will be made publicly available on the Short Read Archive and Genbank under the BioProect accession no. PRJNA1010140.

8 Conflicts of Interest

The authors declare no conflict of interest, financial or otherwise.

References

Águeda, Beatriz et al. (Oct. 2008). "Mycorrhizal synthesis between Boletus edulis species complex and rockroses (Cistus sp.)" en. In: *Mycorrhiza* 18.8, pp. 443–449. ISSN: 0940-6360, 1432-1890.

Alonge, Michael et al. (Dec. 2022). "Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing". In: *Genome Biology* 23.1, p. 258. ISSN: 1474-760X.

Arora, David (2008). "California Porcini: Three New Taxa, Observations on Their Harvest, and the Tragedy of No Commons". In: *Economic Botany* 62.3, pp. 356–375. ISSN: 0013-0001.

Arora, David and Jonathan Frank (Aug. 2014). "Boletus rubriceps , a new species of porcini from the southwestern USA". en. In: *North American Fungi* 9.0, pp. 1–11. ISSN: 1937-786X.

Ayala, Francisco J. et al. (May 2005). "Chromosome speciation: Humans, Drosophila, and mosquitoes". In: *Proceedings of the National Academy of Sciences* 102.suppl_1, pp. 6535–6542.

Badet, Thomas et al. (Feb. 2020). "A 19-isolate reference-quality global pangenome for the fungal wheat pathogen Zymoseptoria tritici". en. In: *BMC Biology* 18.1, p. 12. ISSN: 1741-7007.

Baldrian, Petr (Oct. 2009). "Ectomycorrhizal fungi and their enzymes in soils: is there enough evidence for their role as facultative soil saprotrophs?" en. In: *Oecologia* 161.4, pp. 657–660. ISSN: 1432-1939.

Bankevich, Anton et al. (May 2012). "SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing". In: *Journal of Computational Biology* 19.5, pp. 455–477.

Bayer, Philipp E. et al. (Aug. 2020). "Plant pan-genomes are the new reference". en. In: *Nature Plants* 6.8, pp. 914–920. ISSN: 2055-0278.

Bending, Gary D. et al. (1995). "The Structure and Function of the Vegetative Mycelium of Ectomycorrhizal Plants. VI. Activities of Nutrient Mobilizing Enzymes in Birch Litter Colonized by Paxillus involutus (Fr.) Fr." In: *The New Phytologist* 130.3, pp. 411–417. ISSN: 0028-646X.

Bendixsen, Devin P et al. (Feb. 2021). "Genomic Evidence of an Ancient East Asian Divergence Event in Wild Saccharomyces cerevisiae". In: *Genome Biology and Evolution* 13.2, evab001. ISSN: 1759-6653.

Branco, Sara et al. (Apr. 2017). "Continental-level population differentiation and environmental adaptation in the mushroom *Suillus brevipes*". en. In: *Molecular Ecology* 26.7, pp. 2063–2076. ISSN: 09621083.

Buchfink, Benjamin et al. (Jan. 2015). "Fast and sensitive protein alignment using DIAMOND". en. In: *Nature Methods* 12.1, pp. 59–60. ISSN: 1548-7105.

Chen, Shifu et al. (Sept. 2018). "fastp: an ultra-fast all-in-one FASTQ preprocessor". en. In: *Bioinformatics* 34.17, pp. i884–i890. ISSN: 1367-4803.

Chen, Ying et al. (Jan. 2021). "Efficient assembly of nanopore reads via highly accurate and intact error correction". en. In: *Nature Communications* 12.1, p. 60. ISSN: 2041-1723.

Danecek, Petr et al. (Aug. 2011). "The variant call format and VCFtools". In: *Bioinformatics* 27.15, pp. 2156–2158. ISSN: 1367-4803.

Dentinger, Bryn T. M. et al. (Dec. 2010). "Molecular phylogenetics of porcini mushrooms (Boletus section Boletus)". In: *Molecular Phylogenetics and Evolution* 57.3, pp. 1276–1292. ISSN: 1055-7903.

Ellison, C. E. et al. (Feb. 2011). "Population genomics and local adaptation in wild isolates of a model microbial eukaryote". en. In: *Proceedings of the National Academy of Sciences* 108.7, pp. 2831–2836. ISSN: 0027-8424, 1091-6490.

Emms, David M. et al. (Nov. 2019). "OrthoFinder: phylogenetic orthology inference for comparative genomics". In: *Genome Biology* 20.1, p. 238. ISSN: 1474-760X.

Feurtey, Alice et al. (2018). "Interspecific Gene Exchange as a Driver of Adaptive Evolution in Fungi". In: *Annual Review of Microbiology* 72.1, pp. 377–398.

Finn, Robert D. et al. (July 2011). "HMMER web server: interactive sequence similarity searching". In: *Nucleic Acids Research* 39.Web Server issue, W29–W37. ISSN: 0305-1048.

Frichot, Eric et al. (Apr. 2014). "Fast and Efficient Estimation of Individual Ancestry Coefficients". In: *Genetics* 196.4, pp. 973–983. ISSN: 0016-6731.

Giraud, T et al. (Oct. 2012). "The tempo and modes of evolution of reproductive isolation in fungi". en. In: *Heredity* 109.4, pp. 204–214. ISSN: 0018-067X, 1365-2540.

Gladieux, Pierre et al. (2014). "Fungal evolutionary genomics provides insight into the mechanisms of adaptive divergence in eukaryotes". en. In: *Molecular Ecology* 23.4, pp. 753–773. ISSN: 1365-294X.

Goodwin, Stephen B. et al. (June 2011). "Finished Genome of the Fungal Wheat Pathogen Mycosphaerella graminicola Reveals Dispensome Structure, Chromosome Plasticity, and Stealth Pathogenesis". en. In: *PLOS Genetics* 7.6, e1002070. ISSN: 1553-7404.

Grigoriev, Igor V. et al. (Jan. 2014). "MycoCosm portal: gearing up for 1000 fungal genomes". en. In: *Nucleic Acids Research* 42.D1, pp. D699–D704. ISSN: 0305-1048, 1362-4962.

Hämälä, Tuomas et al. (Aug. 2021). "Genomic structural variants constrain and facilitate adaptation in natural populations of Theobroma cacao, the chocolate tree". In: *Proceedings of the National Academy of Sciences* 118.35, e2102914118.

Hane, James K. et al. (May 2011). "A novel mode of chromosomal evolution peculiar to filamentous Ascomycete fungi". In: *Genome Biology* 12.5, R45. ISSN: 1474-760X.

Hirsch, Candice N. et al. (Jan. 2014). "Insights into the Maize Pan-Genome and Pan-Transcriptome". In: *The Plant Cell* 26.1, pp. 121–135. ISSN: 1040-4651.

Hoang, Diep Thi et al. (Feb. 2018). "UFBoot2: Improving the Ultrafast Bootstrap Approximation". In: *Molecular Biology and Evolution* 35.2, pp. 518–522. ISSN: 0737-4038.

Hyun, Jason C. et al. (Jan. 2022). "Comparative pangenomics: analysis of 12 microbial pathogen pangenomes reveals conserved global structures of genetic and functional diversity". In: *BMC Genomics* 23.1, p. 7. ISSN: 1471-2164.

Inderbitzin, Patrik et al. (Mar. 2011). "The Ascomycete Verticillium longisporum Is a Hybrid and a Plant Pathogen with an Expanded Host Range". en. In: *PLOS ONE* 6.3, e18260. ISSN: 1932-6203.

Ishida, Takahide A. et al. (2007). "Host effects on ectomycorrhizal fungal communities: insight from eight host species in mixed conifer–broadleaf forests". en. In: *New Phytologist* 174.2, pp. 430–440. ISSN: 1469-8137.

Jenny, Hans et al. (Dec. 1949). "COMPARATIVE STUDY OF DECOMPOSITION RATES OF ORGANIC MATTER IN TEMPERATE AND TROPICAL REGIONS". en-US. In: *Soil Science* 68.6, p. 419. ISSN: 0038-075X.

Kalyaanamoorthy, Subha et al. (June 2017). "ModelFinder: fast model selection for accurate phylogenetic estimates". en. In: *Nature Methods* 14.6, pp. 587–589. ISSN: 1548-7091, 1548-7105.

Kang, Heng et al. (2020). “The small secreted effector protein MiSSP7.6 of Laccaria bicolor is required for the establishment of ectomycorrhizal symbiosis”. en. In: *Environmental Microbiology* 22.4, pp. 1435–1446. ISSN: 1462-2920.

Katoh, Kazutaka et al. (Sept. 2017). “MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization”. en. In: *Briefings in Bioinformatics*.

Kaushik, Abhinav et al. (Mar. 2022). “Pangenome Analysis of the Soilborne Fungal Phytopathogen Rhizoctonia solani and Development of a Comprehensive Web Resource: RsolaniDB”. In: *Frontiers in Microbiology* 13, p. 839524. ISSN: 1664-302X.

Khan, Aamir W. et al. (Feb. 2020). “Super-Pangenome by Integrating the Wild Side of a Species for Accelerated Crop Improvement”. In: *Trends in Plant Science* 25.2, pp. 148–158. ISSN: 1360-1385.

Kim, Daehwan et al. (Aug. 2019). “Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype”. en. In: *Nature Biotechnology* 37.8, pp. 907–915. ISSN: 1546-1696.

Kohler, Annegret et al. (Apr. 2015). “Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists”. en. In: *Nature Genetics* 47.4, pp. 410–415. ISSN: 1546-1718.

Kolmogorov, Mikhail et al. (May 2019). “Assembly of long, error-prone reads using repeat graphs”. en. In: *Nature Biotechnology* 37.5, pp. 540–546. ISSN: 1546-1696.

Koren, Sergey et al. (Mar. 2017). “Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation”. en. In: *Genome Research*, gr.215087.116. ISSN: 1088-9051, 1549-5469.

Kuikka, Karita et al. (2003). “Severe Defoliation of Scots Pine Reduces Reproductive Investment by Ectomycorrhizal Symbionts”. en. In: *Ecology* 84.8, pp. 2051–2061. ISSN: 1939-9170.

Kummel, Miroslav et al. (July 2011). “Altering light availability to the plant host determined the identity of the dominant ectomycorrhizal fungal partners and mediated mycorrhizal effects on plant growth”. In: *Botany* 89.7, pp. 439–450. ISSN: 1916-2790.

Ladd, J. N. et al. (1985). “Decomposition of plant material in Australian soils. III. Residual organic and microbial biomass C and N from isotope-labelled legume material and soil organic matter, decomposing under field conditions”. en. In: *Soil Research* 23.4, pp. 603–611. ISSN: 1838-6768.

Langmead, Ben et al. (Mar. 2012). "Fast gapped-read alignment with Bowtie 2". eng. In: *Nature Methods* 9.4, pp. 357–359. ISSN: 1548-7105.

Lindahl, Björn D. et al. (2015). "Ectomycorrhizal fungi – potential organic matter decomposers, yet not saprotrophs". en. In: *New Phytologist* 205.4, pp. 1443–1447. ISSN: 1469-8137.

Liu, Hailin et al. (Mar. 2021). "SMARTdenovo: a de novo assembler using long noisy reads". en. In: *Gigabyte* 2021, pp. 1–9. ISSN: 2709-4715.

Liu, Yucheng et al. (July 2020). "Pan-Genome of Wild and Cultivated Soybeans". en. In: *Cell* 182.1, 162–176.e13. ISSN: 0092-8674.

Lofgren, Lotus A., Nhu H. Nguyen, et al. (2021). "Comparative genomics reveals dynamic genome evolution in host specialist ectomycorrhizal fungi". en. In: *New Phytologist* 230.2, pp. 774–792. ISSN: 1469-8137.

Lofgren, Lotus A., Brandon S. Ross, et al. (Nov. 2022). "The pan-genome of Aspergillus fumigatus provides a high-resolution view of its population structure revealing high levels of lineage-specific diversity driven by recombination". en. In: *PLOS Biology* 20.11, e3001890. ISSN: 1545-7885.

Maan, Martine E. et al. (2011). "Ecology, sexual selection and speciation". en. In: *Ecology Letters* 14.6, pp. 591–602. ISSN: 1461-0248.

McCarthy, Charley G. P. et al. (Feb. 2019). "Pan-genome analyses of model fungal species". In: *Microbial Genomics* 5.2, e000243. ISSN: 2057-5858.

Min, Byoungnam et al. (Jan. 2020). "Unusual genome expansion and transcription suppression in ectomycorrhizal Tricholoma matsutake by insertions of transposable elements". en. In: *PLOS ONE* 15.1, e0227923. ISSN: 1932-6203.

Minh, Bui Quang et al. (May 2020). "IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era". In: *Molecular Biology and Evolution* 37.5, pp. 1530–1534. ISSN: 0737-4038.

Moore, Jamie C. et al. (Mar. 2011). "Sexual selection in plants". en. In: *Current Biology* 21.5, R176–R182. ISSN: 0960-9822.

Morin, Scott J. et al. (Jan. 2017). "Translocations, inversions and other chromosome rearrangements". en. In: *Fertility and Sterility* 107.1, pp. 19–26. ISSN: 0015-0282.

Navarro, A. et al. (June 1997). "Recombination and gene flux caused by gene conversion and crossing over in inversion heterokaryotypes". eng. In: *Genetics* 146.2, pp. 695–709. ISSN: 0016-6731.

NextDenovo (Mar. 2023).

Nguyen, Trang et al. (Nov. 2019). "Contribution of different catalytic types of peptidases to soil proteolytic activity". en. In: *Soil Biology and Biochemistry* 138, p. 107578. ISSN: 0038-0717.

O'Donnell, Samuel et al. (May 2020). "MUM&Co: accurate detection of all SV types through whole-genome alignment". In: *Bioinformatics* 36.10, pp. 3242–3243. ISSN: 1367-4803.

Ou, Shujun et al. (Dec. 2019). "Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline". In: *Genome Biology* 20.1, p. 275. ISSN: 1474-760X.

Palmer, Jon (Oct. 2022). *nextgenusfs/funannotate*.

Passer, Andrew Ryan et al. (June 2019). "Genetic and Genomic Analyses Reveal Boundaries between Species Closely Related to Cryptococcus Pathogens". In: *mBio* 10.3, 10.1128/mbio.00764–19.

Pellegrin, Clément et al. (2019). "Laccaria bicolor MiSSP8 is a small-secreted protein decisive for the establishment of the ectomycorrhizal symbiosis". en. In: *Environmental Microbiology* 21.10, pp. 3765–3779. ISSN: 1462-2920.

Plett, Jonathan M et al. (Mar. 2023). "Speciation Underpinned by Unexpected Molecular Diversity in the Mycorrhizal Fungal Genus Pisolithus". In: *Molecular Biology and Evolution* 40.3, msad045. ISSN: 1537-1719.

Plett, Jonathan M. et al. (July 2011). "A secreted effector protein of Laccaria bicolor is required for symbiosis development". eng. In: *Current biology: CB* 21.14, pp. 1197–1203. ISSN: 1879-0445.

Ravinet, Mark et al. (2021). "Patterns of genomic divergence and introgression between Japanese stickleback species with overlapping breeding habitats". en. In: *Journal of Evolutionary Biology* 34.1, pp. 114–127. ISSN: 1420-9101.

Rawlings, Neil D et al. (Jan. 2018). "The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database". In: *Nucleic Acids Research* 46.D1, pp. D624–D632. ISSN: 0305-1048.

Rieseberg, L. H. (July 2001). "Chromosomal rearrangements and speciation". eng. In: *Trends in Ecology & Evolution* 16.7, pp. 351–358. ISSN: 1872-8383.

Samarasinghe, Himeshi et al. (Jan. 2020). "Hybridization Facilitates Adaptive Evolution in Two Major Fungal Pathogens". In: *Genes* 11.1, p. 101. ISSN: 2073-4425.

Samsatly, Jamil et al. (2020). "Vitamin B6 Is Under a Tight Balance During Disease Development by Rhizoctonia solani on Different Cultivars of Potato and on Arabidopsis thaliana Mutants". In: *Frontiers in Plant Science* 11. ISSN: 1664-462X.

Seehausen, Ole et al. (Mar. 2014). "Genomics and the origin of species". en. In: *Nature Reviews Genetics* 15.3, pp. 176–192. ISSN: 1471-0064.

Semenova, Tatyana A. et al. (May 2017). "Extracellular peptidases as possible markers of fungal ecology". en. In: *Applied Soil Ecology* 113, pp. 1–10. ISSN: 0929-1393.

Shah, Firoz et al. (2013). "The molecular components of the extracellular protein-degradation pathways of the ectomycorrhizal fungus Paxillus involutus". en. In: *New Phytologist* 200.3, pp. 875–887. ISSN: 1469-8137.

Sherman, Rachel M. et al. (Apr. 2020). "Pan-genomics in the human genome era". en. In: *Nature Reviews Genetics* 21.4, pp. 243–254. ISSN: 1471-0064.

Shoji, Jun-ya et al. (Apr. 2015). "Vegetative Hyphal Fusion and Subsequent Nuclear Behavior in Epichloë Grass Endophytes". en. In: *PLOS ONE* 10.4, e0121875. ISSN: 1932-6203.

Smith, Alexander Hanchett (2005). *The boletes of Michigan, by Alexander H. Smith and Harry D. Thiers.*

Sperschneider, Jana et al. (Feb. 2022). "EffectorP 3.0: Prediction of Apoplastic and Cytoplasmic Effectors in Fungi and Oomycetes". In: *Molecular Plant-Microbe Interactions®* 35.2, pp. 146–156. ISSN: 0894-0282.

Steensels, Jan et al. (Aug. 2021). "Interspecific hybridization as a driver of fungal evolution and adaptation". en. In: *Nature Reviews Microbiology* 19.8, pp. 485–500. ISSN: 1740-1534.

Steidinger, B. S. et al. (May 2019). "Climatic controls of decomposition drive the global biogeography of forest-tree symbioses". en. In: *Nature* 569.7756, pp. 404–408. ISSN: 1476-4687.

Talbot, J. M. et al. (2008). "Decomposers in disguise: mycorrhizal fungi as regulators of soil C dynamics in ecosystems under global change". en. In: *Functional Ecology* 22.6, pp. 955–963. ISSN: 1365-2435.

Talbot, Jennifer M. et al. (Feb. 2013). "Independent roles of ectomycorrhizal and saprotrophic communities in soil organic matter decomposition". en. In: *Soil Biology and Biochemistry* 57, pp. 282–291. ISSN: 0038-0717.

Tedersoo, Leho, Mohammad Bahram, et al. (2012). "Towards global patterns in the diversity and community structure of ectomycorrhizal fungi". en. In: *Molecular Ecology* 21.17, pp. 4160–4170. ISSN: 1365-294X.

Tedersoo, Leho and Kazuhide Nara (2010). "General Latitudinal Gradient of Biodiversity Is Reversed in Ectomycorrhizal Fungi". In: *The New Phytologist* 185.2, pp. 351–354. ISSN: 0028-646X.

Teufel, Felix et al. (2022). "SignalP 6.0 predicts all five types of signal peptides using protein language models". In: *Nature Biotechnology* 40.7, pp. 1023–1025. ISSN: 1087-0156.

Tigano, Anna et al. (2016). "Genomics of local adaptation with gene flow". en. In: *Molecular Ecology* 25.10, pp. 2144–2164. ISSN: 1365-294X.

Tremble, Keaton, J.i. Hoffman, et al. (Oct. 2022). "Contrasting continental patterns of adaptive population divergence in the holarctic ectomycorrhizal fungus Boletus edulis". en. In: *New Phytologist* n/a.n/a. ISSN: 1469-8137.

Tremble, Keaton, Laura M. Suz, et al. (Oct. 2019). "Lost in Translation: population genomics and long-read sequencing reveals relaxation of concerted evolution of the ribosomal DNA cistron". en. In: *bioRxiv*, p. 811216.

— (July 2020). "Lost in translation: Population genomics and long-read sequencing reveals relaxation of concerted evolution of the ribosomal DNA cistron". en. In: *Molecular Phylogenetics and Evolution* 148, p. 106804. ISSN: 1055-7903.

Uhse, Simon et al. (June 2018). "Effectors of plant-colonizing fungi and beyond". en. In: *PLOS Pathogens* 14.6, e1006992. ISSN: 1553-7374.

Vaario, Lu-Min et al. (Jan. 2019). "Tricholoma matsutake may take more nitrogen in the organic form than other ectomycorrhizal fungi for its sporocarp development: the isotopic evidence". en. In: *Mycorrhiza* 29.1, pp. 51–59. ISSN: 1432-1890.

Vernikos, G. S. (2020). "A Review of Pangenome Tools and Recent Studies". en. In: *The Pangenome: Diversity, Dynamics and Evolution of Genomes*. Ed. by Hervé Tettelin et al. Cham: Springer International Publishing, pp. 89–112. ISBN: 978-3-030-38281-0.

Walker, Bruce J. et al. (Nov. 2014). "Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement". en. In: *PLOS ONE* 9.11, e112963. ISSN: 1932-6203.

Wang, Q.B. et al. (June 2005). "Boletus reticuloceps, a new combination for Aureoboletus reticuloceps". In: *Sydowia* 57, pp. 131–136.

Wang, Xinfeng et al. (Dec. 2022). "Extensive gene flow in secondary sympatry after allopatric speciation". In: *National Science Review* 9.12, nwac280. ISSN: 2095-5138.

Waterhouse, Robert M. et al. (Mar. 2018). "BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics". en. In: *Molecular Biology and Evolution* 35.3, pp. 543–548. ISSN: 0737-4038.

Wellenreuther, Maren et al. (2019). "Going beyond SNPs: The role of structural genomic variants in adaptive evolution and species diversification". en. In: *Molecular Ecology* 28.6, pp. 1203–1209. ISSN: 1365-294X.

West-Eberhard, Mary Jane (June 1983). "Sexual Selection, Social Competition, and Speciation". In: *The Quarterly Review of Biology* 58.2, pp. 155–183. ISSN: 0033-5770.

Wong-Bajracharya, Johanna et al. (Jan. 2022). "The ectomycorrhizal fungus Pisolithus microcarpus encodes a microRNA involved in cross-kingdom gene silencing during symbiosis". In: *Proceedings of the National Academy of Sciences* 119.3, e2103527119.

Wood, Derrick E. et al. (Nov. 2019). "Improved metagenomic analysis with Kraken 2". In: *Genome Biology* 20.1, p. 257. ISSN: 1474-760X.

Wu, Gang et al. (2022). "Evolutionary innovations through gain and loss of genes in the ectomycorrhizal Boletales". en. In: *New Phytologist* 233.3, pp. 1383–1400. ISSN: 1469-8137.

Xu, Jing et al. (Apr. 2020). "eCAMI: simultaneous classification and motif identification for enzyme annotation". In: *Bioinformatics* 36.7, pp. 2068–2075. ISSN: 1367-4803.

Zhang, Chao et al. (May 2018). "ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees". en. In: *BMC Bioinformatics* 19.S6. ISSN: 1471-2105.

Zhang, Han et al. (July 2018). "dbCAN2: a meta server for automated carbohydrate-active enzyme annotation". In: *Nucleic Acids Research* 46.W1, W95–W101. ISSN: 0305-1048.

Zhang, Linyi et al. (July 2021). "How Important Are Structural Variants for Speciation?" en. In: *Genes* 12.7, p. 1084. ISSN: 2073-4425.

Zimin, Aleksey V. et al. (Nov. 2013). "The MaSuRCA genome assembler". en. In: *Bioinformatics* 29.21, pp. 2669–2677. ISSN: 1367-4803.

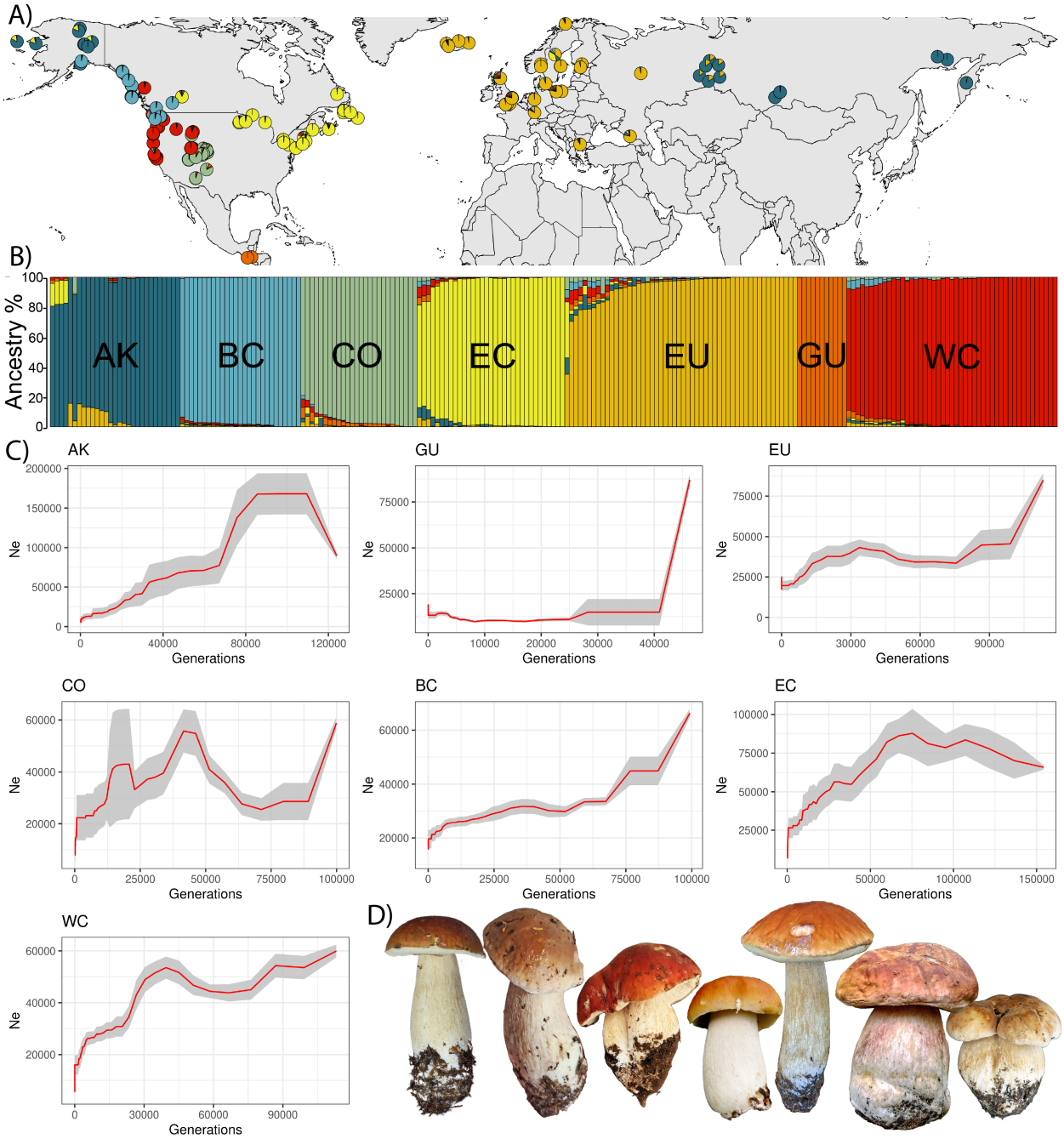
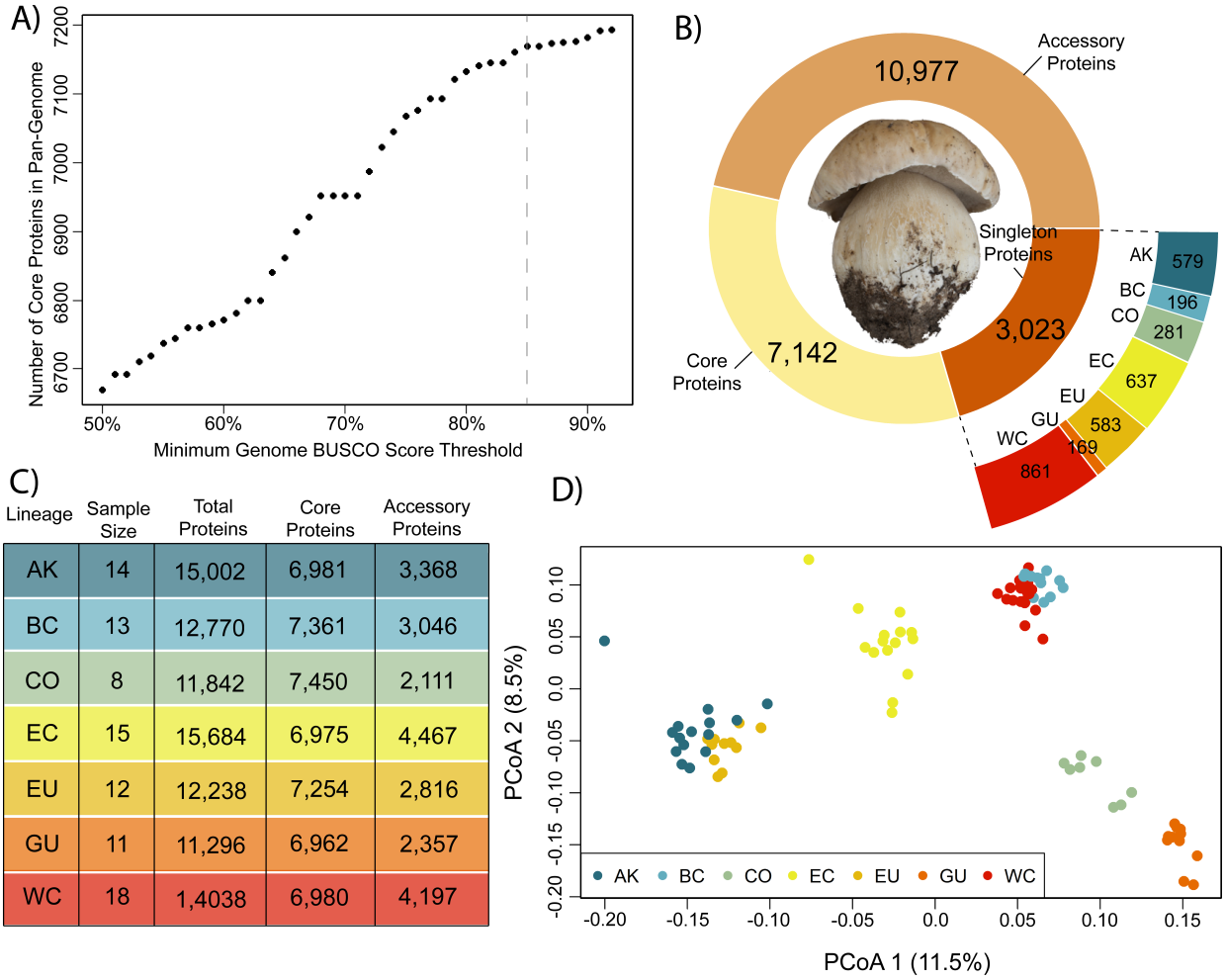


Figure 1: Results of admixture and demographic analysis of *B. edulis*. A) Distribution of estimated admixture across the global distribution of *B. edulis*. B) Barchart of admixture proportions per sample per lineage. C) Estimated changes in historical effective population sizes (N_e) per lineage. Grey regions indicate 95% confidence intervals. D) Examples of the phenotypic diversity found within *B. edulis*. From left to right: YSU-5404 (AK), KST15 (BC), BD942 (CO), BD294 (EC), KST39 (GU), UT567 (WC), BD747 (WC), all photos are B. Dentinger, except YSU-5404 (AK) and KST15 (BC). It is important to note that phenotypes do not cleanly segregate into lineages.



Lineage	Sample Size	Total Proteins	Core Proteins	Accessory Proteins
AK	14	15,002	6,981	3,368
BC	13	12,770	7,361	3,046
CO	8	11,842	7,450	2,111
EC	15	15,684	6,975	4,467
EU	12	12,238	7,254	2,816
GU	11	11,296	6,962	2,357
WC	18	1,4038	6,980	4,197

Figure 2: The global pan-genome of *B. edulis*. A) The impact of minimum assembly quality threshold on core protein abundance. The solid grey line represents the quality threshold used (85% minimum BUSCO score) for the final pan-genome analysis. B) The abundance of core (>=95% presence), accessory (2-94% presence), and singleton (N=1) within the *B. edulis* pan-genome. C) The abundance of core and accessory proteins found within each individual lineage of *B. edulis*. D) Principle component analysis (PCoA) of 92 *B. edulis* genomes based on accessory protein profiles.

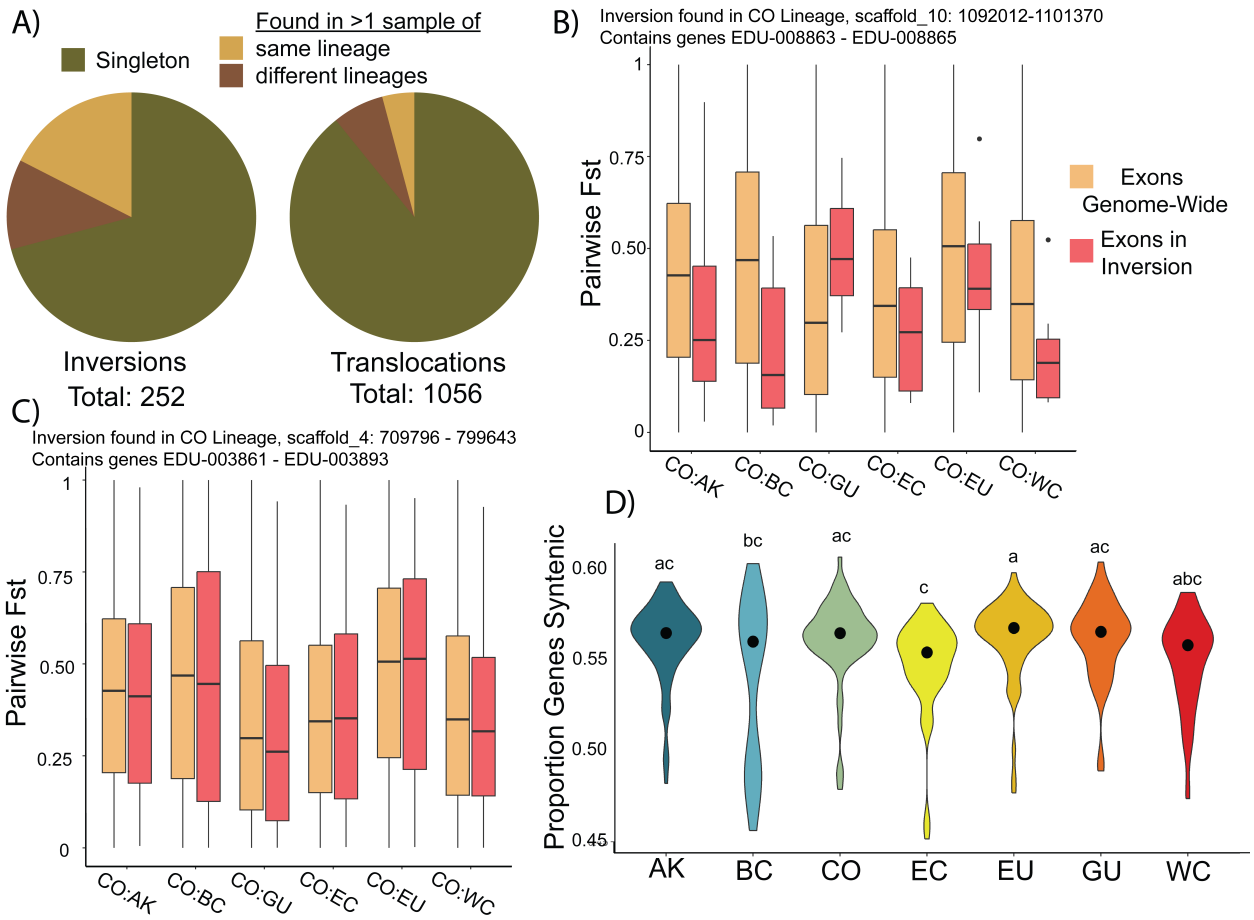


Figure 3: Analysis of large genomic structural variation within *B. edulis*. A) Abundance and distribution of large insertions (left) and translocations (right) within 22 *B. edulis* reference genomes. B) Pairwise F(st) divergence analysis of protein coding regions with an insertion that is potentially adaptive in the CO lineage. This insertion is an example of a structural variant that potentially contributes to local adaptation, as coding regions within this variant are less divergent than the rest of the genome for most comparisons, but are highly diverged when compared to the sister lineages of CO and GU. C) Pairwise F(st) divergence analysis of protein coding regions with an insertion that is potentially adaptive in the CO lineage. This insertion is an example of a structural variant that likely does not contribute greatly to divergence, as protein coding regions are not more or less diverged than the rest of the genome. D) The proportion of genes with syntenic flanking regions, assessed in pairwise combinations. Leters indicate statistical significance ($P < 0.05$, ANOVA)

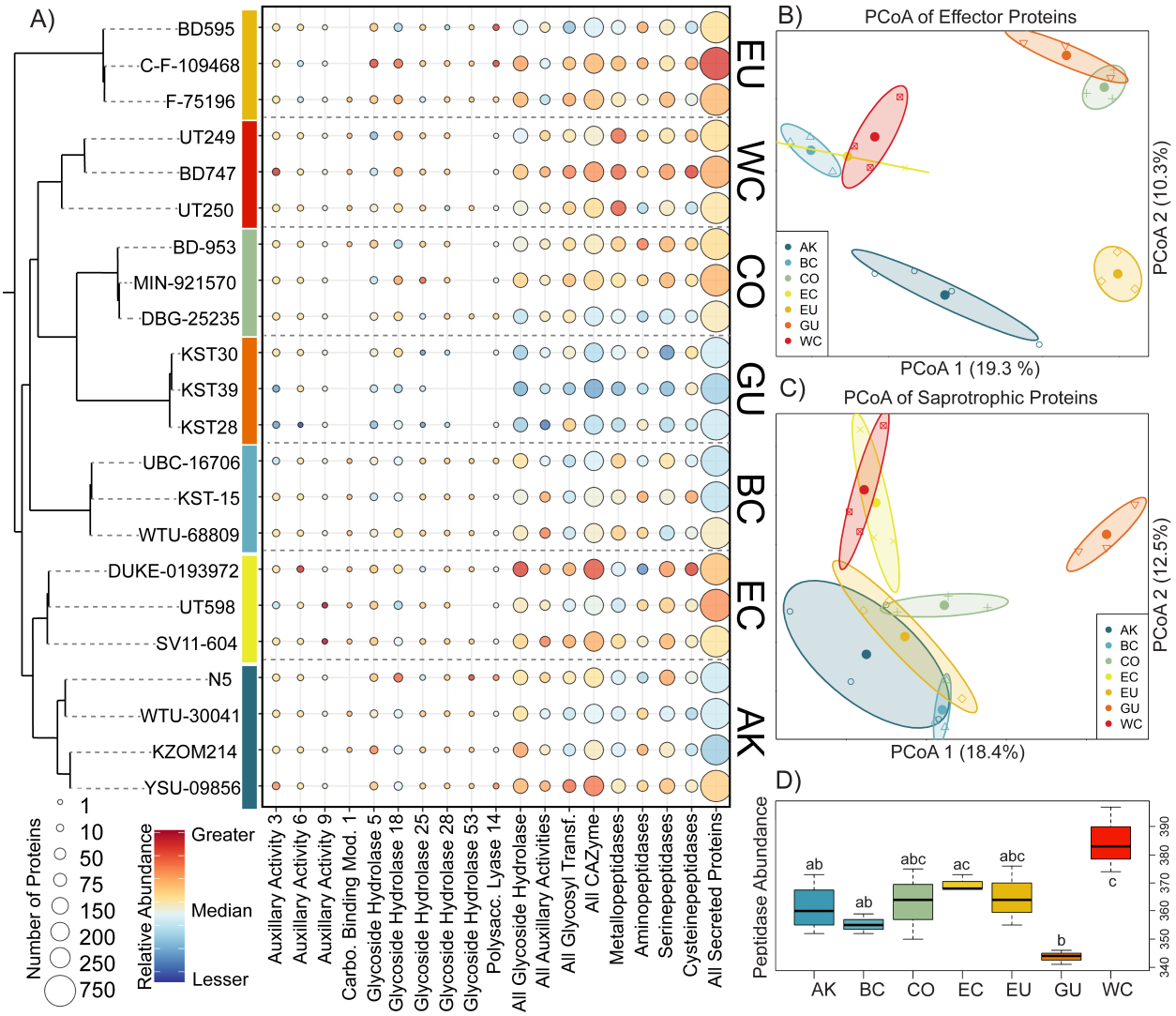


Figure 4: Results of comparative genomic analysis of *B. edulis*. A) Abundance of CAZymes, peptidases, and secreted proteins across the seven lineages of *B. edulis*. The size of the bubbles indicates the number of proteins within each protein family. The lack of a bubble indicates the complete loss of an enzyme family. The color of the bubble indicates the relative abundance within *B. edulis*, where blue indicates relative contractions and red indicates relative expansions. B) Results of the principle component analysis (PCoA) of effector protein repertoires. Elipses indicate 95% confidence intervals. C) Results of the PCoA of saprotrophic protein repertoires. Elipses indicate 95% confidence intervals, D) Abundance of peptidases enzymes among the seven lineages of *B. edulis*. Letters indicate statistical significance (ANOVA, $P < 0.05$).

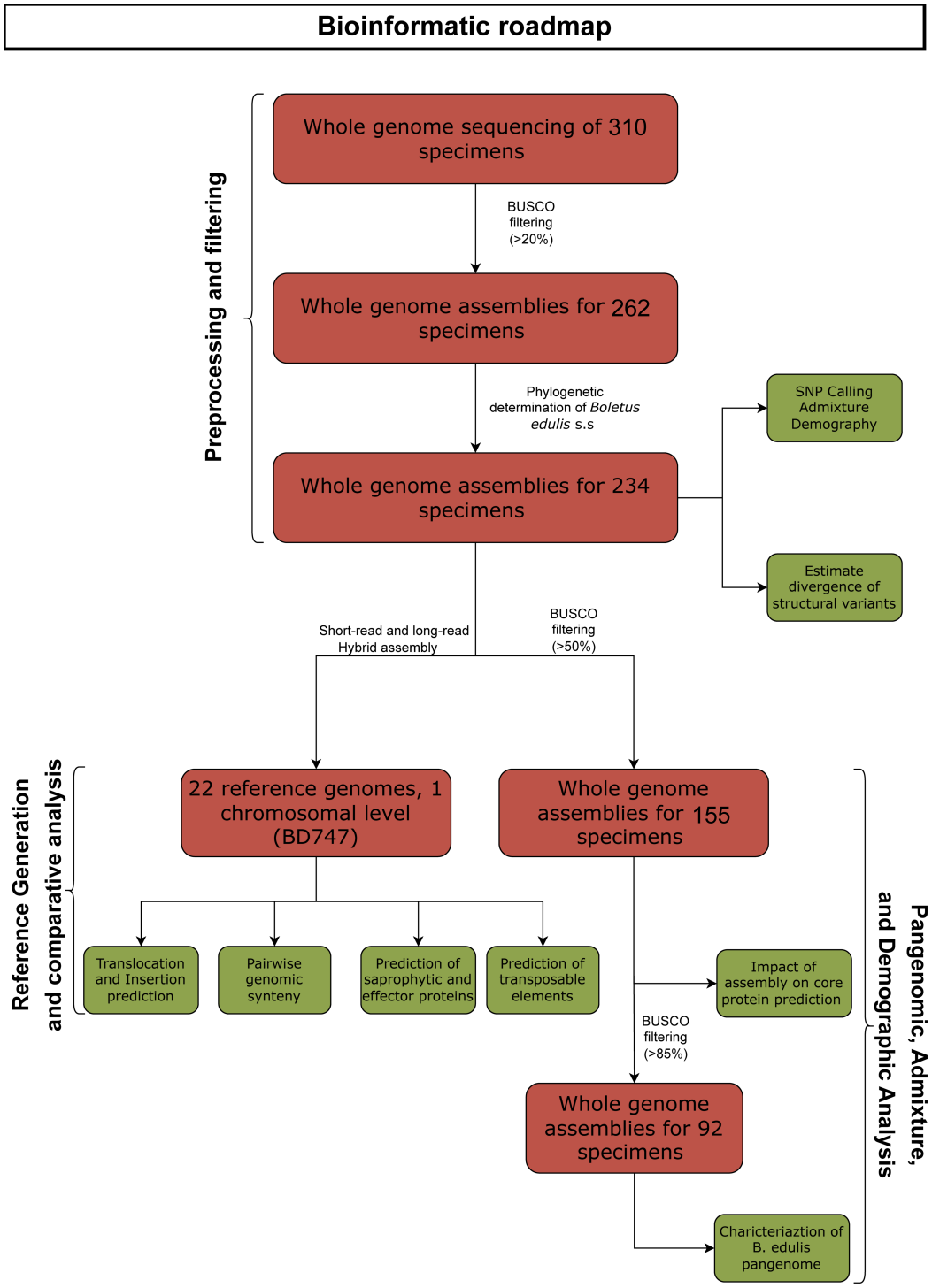


Fig. S1: Flowchart of sample filtering steps and analyses used. Red boxes represent steps where the number of samples was reduced (filtering stages), and green boxes represent analyses performed.

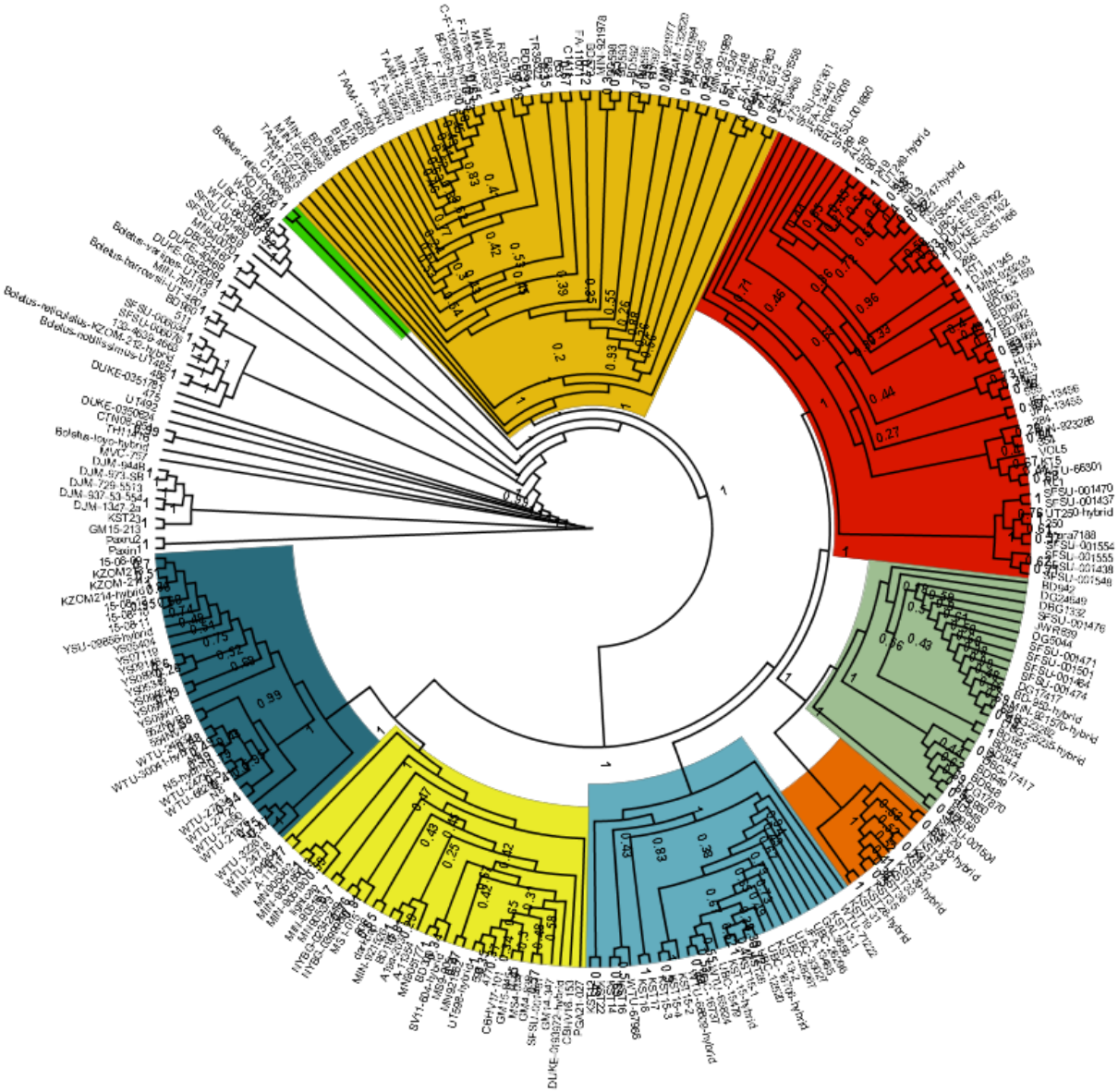


Fig. S2: Summary coalescent phylogeny of *B. edulis*. The sister taxon to *B. edulis*, *B. reticuloceps*, is highlighted in bright green. Monophyloetic lineages within *B. edulis* are highlighted as follows, EU = gold, WC = red, CO = light green, GU = orange, BC = light blue, EC = yellow, AK = dark blue.

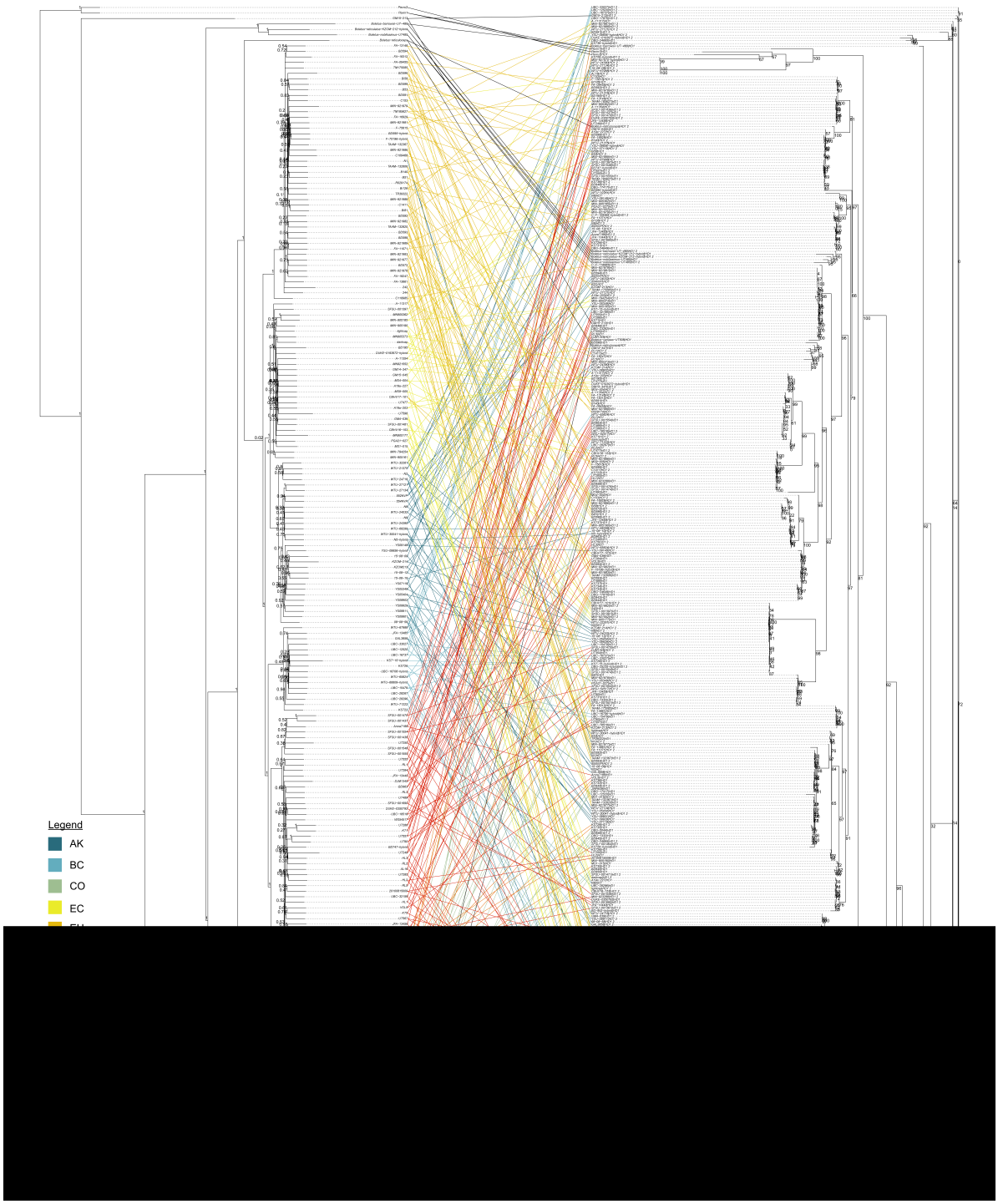


Fig. S3: Cophylogenetic comparison of mating type loci HD1 orthologs and *B. edilus* lineages: Left, summary coalescent species tree of *B. edulis* with node labels corresponding to quartet scores. Right, maximum likelihood tree of recovered HD1 genes, with node labels corresponding to bootstrap support. Connecting lines are colored to correspond with each distinct *B. edulis lineage*.

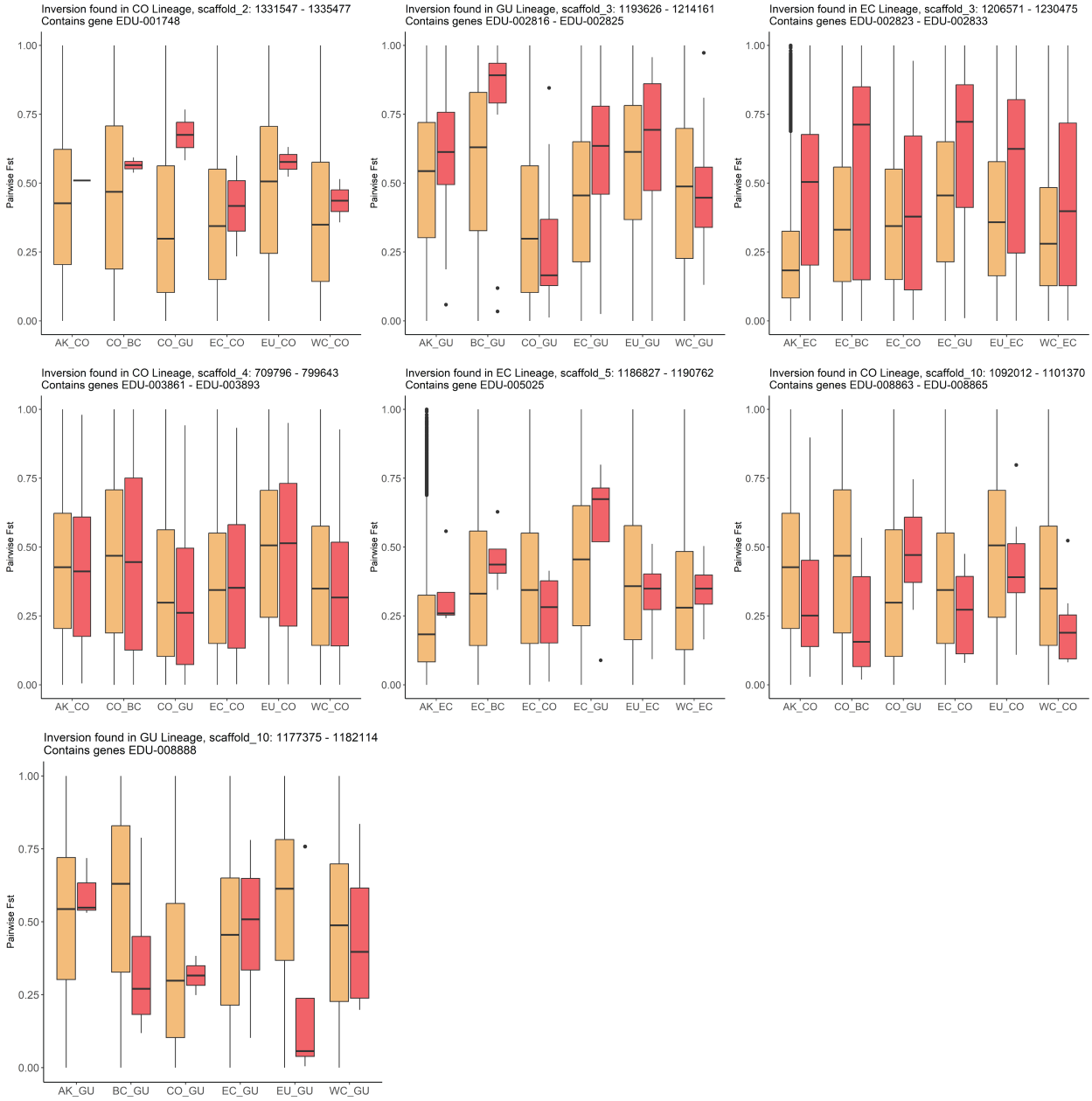


Fig. S4: Comparison of nucleotide divergence F(st) of exons within structural potentially adaptive insertions and exons across the entire genome. The tan bars represent average F(st) values of exons across the genome calculated on a pairwise basis. The pink bars represent F(st) values of exons located within the specific insertion (see chart title for information on structural variant).

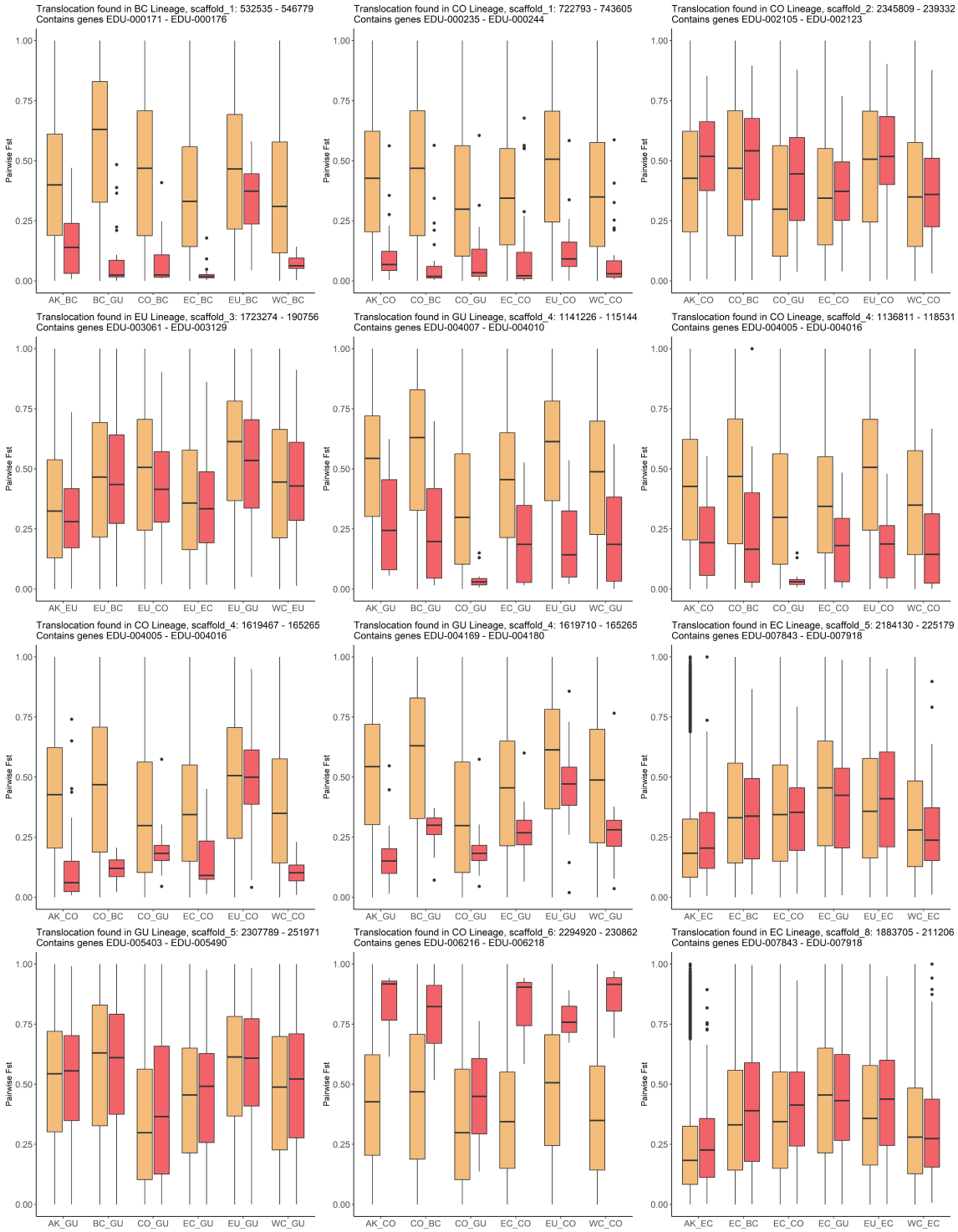


Fig. S5: Comparison of nucleotide divergence F(st) of exons within structural potentially adaptive translocations and exons across the entire genome. The tan bars represent average F(st) values of exons across the genome calculated on a pairwise basis. The pink bars represent F(st) values of exons located within the specific insertion (see chart title for information on structural variant).

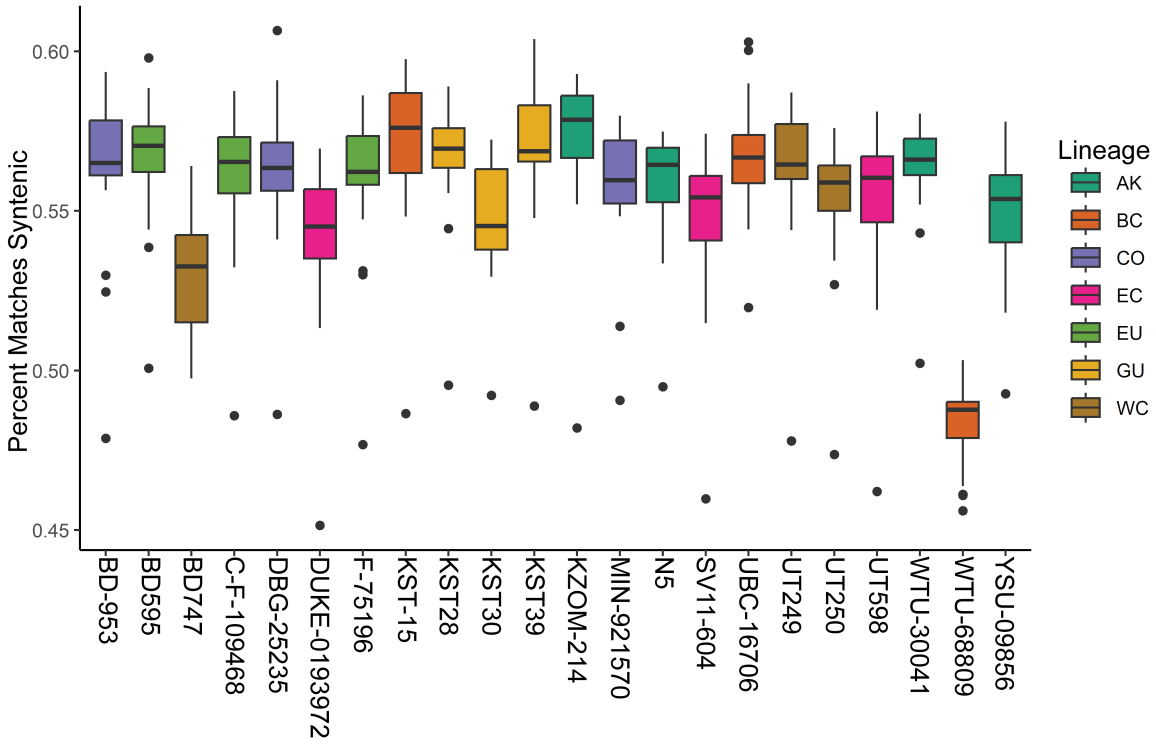


Fig. S6: Proportion of genes with syntenic flanking regions per reference-quality genome.

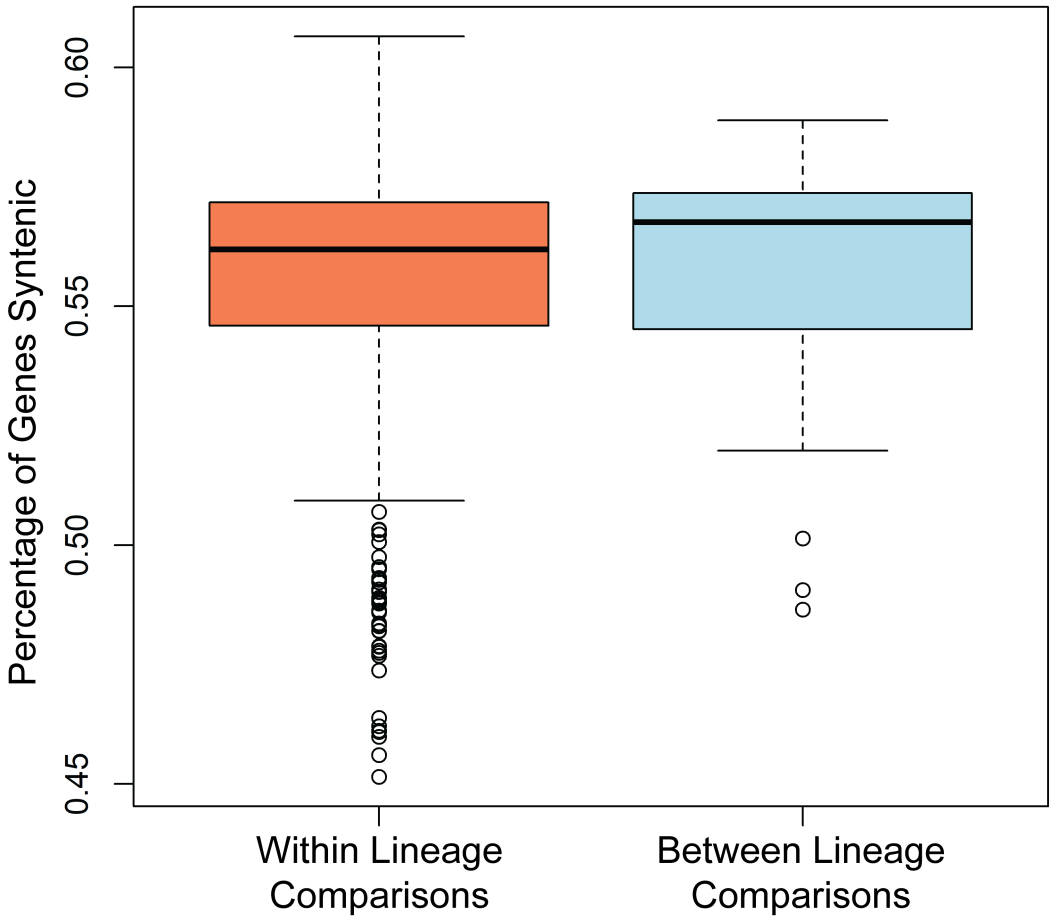


Fig. S7: Proportion of genes with syntenic flanking regions within lineages and between lineages.

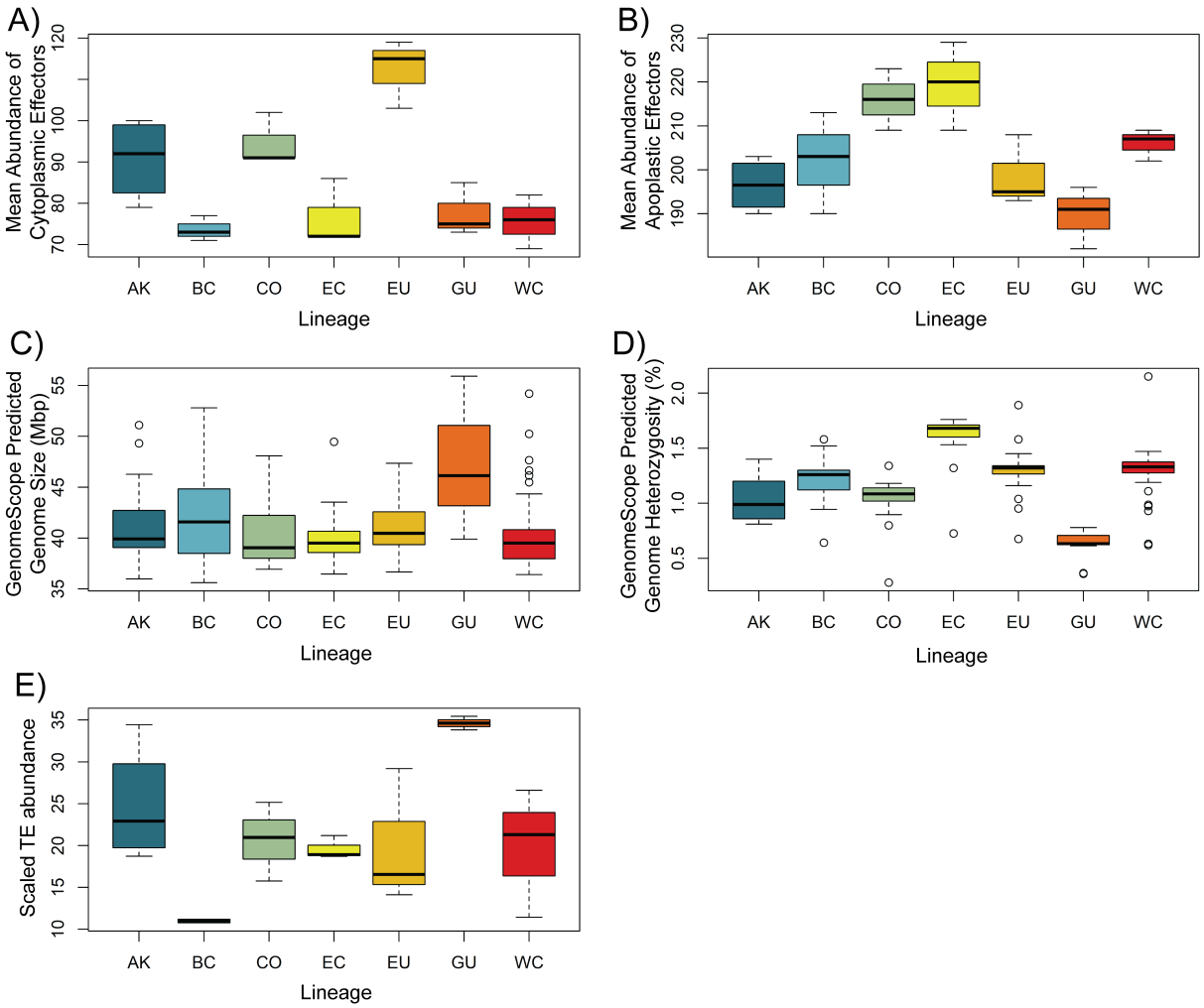


Fig. S8: Boxplots comparaing genomic characteristics of 22 reference genomes in *B. edulis*. A) The average abundance of predicted cytoplasmic effector proteins within lineages. B) The average abundance of predicted appoplastic effector proteins within lineages. C) The average genome size per lineage as modeled with GenomeScope using raw Illumina sequencing reads. D) The average genome heterozygosity percentage per lineage as modeled with GenomeScope using raw Illumina sequencing reads. E) The abundance of genomic transposable element content within lienages of *B. edulis*. The abundance of TEs is scaled by the number of scaffolds within each reference assembly.

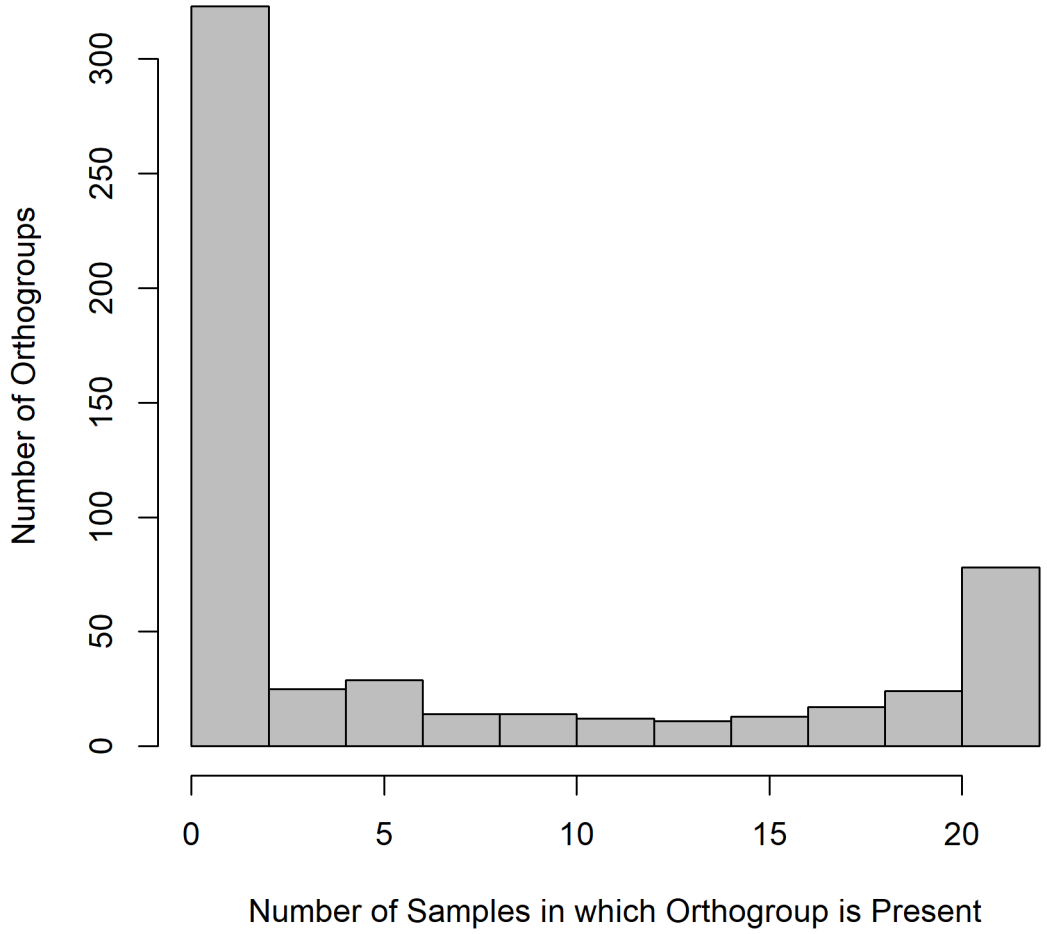


Fig. S9: Histogram of predicted effector protein abundance in 22 *B. edulis* reference genomes.

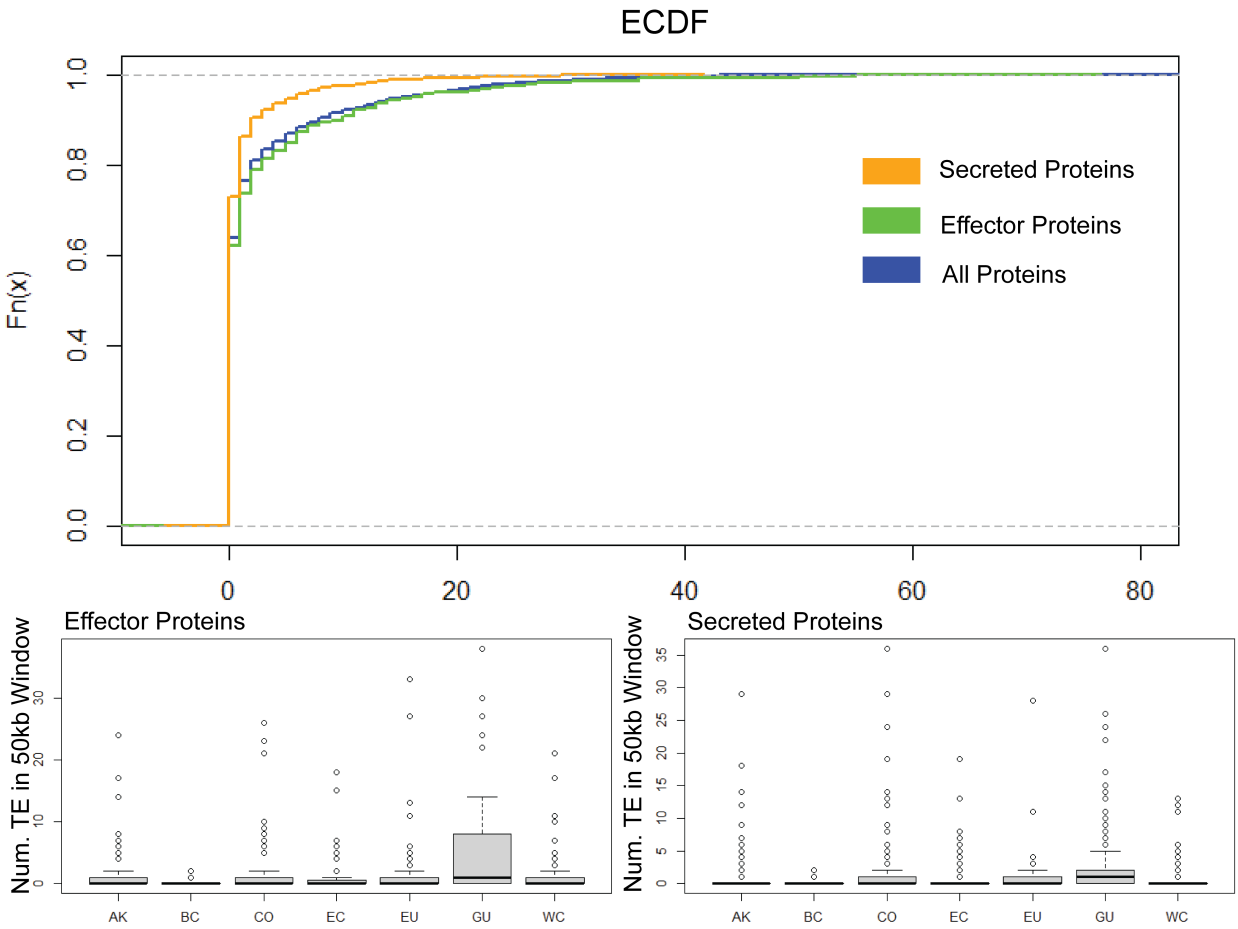


Fig. S10: Top) Distribution of of transposable element density around all proteins (blue line), effector proteins (green line), and secreted proteins (orange line).