Supplementary methods and figures for: Contrasting continental patterns of adaptive population divergence in a holarctic fungus

Keaton Tremble[1,2*], Hoffman, J.I.[3] and Bryn T.M. Dentinger[1,2]

[1]Univeristy of Utah, School of Biological Sciences, Salt Lake City, Utah, 84112.
[2]Natural History Museum of Utah, Salt Lake City, Utah, 84108.
[3]Bielefeld University, Department of Animal Behaviour, Bielefeld, Germany.

*Corresponding author(s). E-mail(s): keaton.tremble@utah.edu;

1 Supplementary Materials and Methods

1.1 Sampling and voucher information

Boletus edulis specimens were gathered from a global distribution using a combination of targeted field collection and through institutional loans from fungaria in North America, Asia, and Europe. Full collection and voucher details are presented in Table S1. No specimen collected prior to 1950 was sampled to maximise DNA library integrity and minimize an effect of time on population stratification. In addition, specimens of the sister group to *B. edulis*, *Boletus reticuloceps*, and two other members of the "porcini" clade, *Boletus variipes* and *Boletus barowsii*, were included as outgroups [1].

1.2 DNA extraction, library preparation, and sequencing

Genomic DNA was extracted from 10mg of hymenophore tissue from each specimen using the Monarch gDNA extraction kit. DNA extract quality was assessed for quality using a NanoDrop 1000 (Thermo Scientific) and fragment integrity using agarose gel electrophoresis. Ssmples were sequenced using a combination of paired-end sequencing on the Illumina MiSeq, HiSeq,

and Novaseq sequencing platforms (Table S2). A highly contiguous reference genome for subsequent variant calling was produced using the MaSuRCA v3.3.1 [2] hybrid-assembly approach using Illumina HiSeq short-read and Oxford Nanopore long-read sequences from the specimen BD747 collected in the Uinta Mountains in Utah, USA. See [3] for full details.

1.3 Genome assembly and phylogenomics

Raw sequencing reads were quality-filtered and adapter-trimmed using fastP v0.20.1 [4] with default settings. Genome assemblies were produced from quality-filtered reads using SPAdes v3.15.0 [5] with five k-mer values (k=77,85,99,111,127). In addition, we included a publicly-available assembly from *Paxillus involutus* [6]retrieved from the JGI Mycocosm Portal [7] as the outgroup. From these assemblies we used exonerate v2.2.0 [8] to extract 702 single-copy orthologs conserved across the Boletales identified by the Boletales MCL clustering in Mycocosm. Retained orthologs were aligned using MAFFT v7.397 [9] with the L-INS-i algorithm, and maximum-likelihood gene-trees were produced using IQ-TREE v2.0.3 [10] with automatic model selection in ModelFinder [11] and ultrafast bootstrapping [12] with 1000 replicates. Gene trees with poor support (BS¡=10%) were removed before generating a summary coalescent species tree using ASTRAL-III v5.7.5 [13]. The "*Boletus edulis*" group was defined by the monophyletic group sharing a most recent common ancestor with *Boletus reticuloceps*, and geographically monphyletic genetic clusters within *B. edulis* were identified as distinct lineages. To date the divergence of the *B. edulis* geographic lineages, we utilized the multiple-species-coalescent model within the Bayesian platform StarBEAST2 [14] using an estimated 100 MYA (95% CI = 70-150 mya) divergence date for the *Boletaceae* [15]. To estimate divergence times using a manageable dataset, a matrix was composed of two arbitrarily selected, but geographically different, representatives from each *B. edulis* lineage, with *B. reticuloceps*, *B. variipes*, and *P. involutus* as outgroups. Twenty of the 702 single-copy genes were randomly selected and putative orthologs of these genes in each genome were aligned in MAFFT, as above. Best-clock models were automatically chosen with bModelTest, site models were set to all-reversible, and an *a priori* initial clock rate of 0.001 was chosen based on estimated mutation rates in related taxa [16]. The MCMC chain was run for $10\hat{7}$ cycles and all parameters reached acceptable convergence based on ESS values ¿ 200. The final tree was produced using Tree-Annotator with an empirically determined burn-in (10%).

1.4 Gene annotation of BD747 reference sequence

To improve gene prediction accuracy and model training, we extracted total RNA from BD747 hymenophore tissue using TRI Reagent® (Sigma-Aldrich) following the manufacturer's instructions. Short-read mRNA enrichment, library prep, and paired-end sequencing was performed by Novogene USA. Full-length cDNA was sequenced on a single Oxford Nanopore Technologies

R9.4 flow cell in a MinION device after mRNA enrichment and library preparation using a Direct cDNA Sequencing Kit (SQK-DCS109) following the manufacturer's protocols. The Funannotate v1.8.6 (funannotate.readthedocs.io) genome annotate wrapper script was used for gene prediction of BD747 according to best practices. In brief, short-read DNA/RNA and long-read cDNA were used to first create a PASA training model, which is utilized to guide Augustus gene-prediction with assembled mRNA transcript evidence. Functional annotation of predicted genes utilizes InterProScan5, Eggnog-Mapper v2.0.8, dbCAN2, SignalP v5.0, and antiSMASH v6.

1.5 Variant calling and SNP filtering

Quality filtered reads were mapped to our *B. edulis* reference BD747, using Bowtie2 v2.2.6 [17]. Variant calling was performed with GATK v4.20 [18] using their best practices for non-model organisms. To retain only high confidence variants, positions with ¿5% missing data, a minimum allele count of 2, and min/max mean read depth of 10 and 75, respectively, were removed with VCFtools v0.1.15 [19]. After this confidence filtering, we then removed any individual with more than 30% missing data to prevent spurious assignments to population due to limited information. Some variant statistical analyses such as PCA and admixture require independent observations, therefore we used PLINK v1.9 [20] to prune out SNPs in high linkage-disequilibrium (r^2 ¿ 0.2, 50KB windows, 10BP step) to create a secondary LD-pruned dataset.

1.6 Population structure analyses

To identify overlap between phylogenomic *B. edulis* lineages and population structure we performed an initial PCA with our LD-pruned dataset using the R package adegenet v2.1.3. To identify sub-population structure of *B. edulis* lineages, PCAs and Neighbor-Joining trees were produced for each lineage. Ancestry admixture analysis was performed with the snmf function of the R package LEA v3.2.0 [21] on the LD-pruned dataset using sites with no missing values, K values 4-8, and 10 runs per K. The run with the lowest cross-entropy for each K was selected for final admixture assessment. Population summary statistics for each lineage were calculated using the non-LD pruned dataset in the R package PopGenome v2.7.5 across the top 100 largest genomic scaffolds. Pairwise F_{st}, and $D_{(xy)}$, were calculated across 10KB non-overlapping windows using the program PIXY (v1.2.2) [22] which facilitates the use of invariant sites to compensate for missing data which is critical for accurate assessments of $D_{(xy)}$. Nucleotide diversity across synonymous and nonsynonymous sites was calculated across all genes on scaffold-1 using the program SNPGenie [23]. To determine if differentiation was limited to non-coding regions, we limited our dataset to include only genic regions and identified the number of highly differentiated genes for each population (F^{st} vs. all ¿ 0.75). To verify the presence of gene flow between lineages/populations we identified the strength of introgression between groups by calculating Patterson's D statistic (ABBA-BABA)

using the Dsuite command line package [24] on all phylogeny-compatible quartets. LD decay of each lineage was assessed using pairwise $R\hat{2}$ values produced with VCFtools across 50,000 Bp windows across the 1.6 Mbp scaffold_1.

1.7 Local adaptation analysis

To identify highly divergent loci that may represent instances of local adaptation, we utilized a two part analysis that identified highly divergent genes containing putative locally adapted SNPs in lineages of *B. edulis*. First, we calculated pairwise F_st and $D(xy)$ statistics with PIXY using only genic regions in our all-sites filtered dataset (Fig S1). We then calculated the average pairwise F_st and $D_(xy)$ for each gene and for each population. Our final dataset of highly divergent loci included genes that were in the top 5% of both F_st and $D_(xy)$. To identify divergent loci that have significant signatures of local adaptation we used the program PCAdapt [25] to identify single putative SNPs within our LD-pruned dataset. We used a K value of K=3 according to proscribed practices as this represented the first plateau identified in a scree plot (Fig. S9), and performed Bonferroni correction (alpha=0.01) to produce a high-confidence dataset. We then used a custom script to identify highly divergent loci that contained SNPs with significant signatures of local adaptation (p¡0.05).

1.8 Isolation by distance/environment and ecological niche modeling

To identify the environmental forces that may be contributing to population and phylogenetic divergence we identified BioClim variables that significantly associated with 1) population clustering based on Principle Component Analysis, 2) putative locally adaptive SNPs using Genetic by Environmental Analysis, and 3) phylogenetic distance between tips across our single-copy orthologue phylogeny. First, we used the R pacakge gdalUtils to identify the 19 BioClim variables at a 30 second scale for each specimen based on their collection location.

To identify signatures of isolation by distance (IBD) and isolation by environment (IBE) we used Multiple Matrix Regression to assess the correlation between geographic, environmental, and genetic distance matrices as well as the interaction between Geographic and Environmental distance. Geographic distances were calculated using GPS coordinates and the dist() function of base R. The 19 Bioclim variables for each individual were plotted using principle component analyses, then Euclidean distances between individuals on the PCA were calculated using the pca2dist function of the R package APE and stored as a final environmental distance matrix. We used two methods to calculate genetic distances: 1) absolute genetic distances using the dist.gene function in ape and 2) first using PCA clustering and then calculating Euclidean distances between individuals, similar to how we calculated environmental distances.

To identify whether current distributions of *B. edulis* populations represent distinct ecological niches we used the Maxent v3.4.1 [26] program to model the potential niche distribution likelihoods of all *B. edulis* lineages based on historic (1970-2000) environmental conditions as determined by the 19 BioClim variables. Default settings were used with 100 replicates and a . In addition, we reran this analysis on a second dataset of the 19 Bioclim variables produced from models of environmental conditions during the last glacial maximum approximately 22,000 years ago.

1.9 Demographic modeling

To understand the demographic history of the *B. edulis* lineages we simulated five different two-population demographic scenarios (Fig 5) and compared these models to two empirical datasets of sister lineages: Westcoast/Colorado and Alaska/Eastcoast. Our specific models are 1) isolation without migration, 2) ancient gene flow followed by isolation, 3) isolation followed by recent gene flow (secondary contact), 4) continuous gene flow after divergence, and 5) differential gene flow as reproductive barriers form (high then low). Empirical datasets consisted of 2-D unfolded site frequency spectra between each lineage pair produced using easySFS. Models were implemented using fastsimcoal2 v2.6.0.3 [27]. Each model was run for 200,000 simulations across 50 iterations and rerun 100 times. We used the Akaike information criterion to determine the best run for each model. To find the best model for each pair of sister taxa, we reran each model using the parameters determined in the best run 100 times and compared the distribution of likelihoods between all five models.

References

[1] Dentinger, B.T.M., Ammirati, J.F., Both, E.E., Desjardin, D.E., Halling, R.E., Henkel, T.W., Moreau, P.-A., Nagasawa, E., Soytong, K., Taylor, A.F., Watling, R., Moncalvo, J.-M., McLaughlin, D.J.: Molecular phylogenetics of porcini mushrooms (Boletus section Boletus). Molecular Phylogenetics and Evolution **57**(3), 1276–1292 (2010). https://doi.org/10.1016/j.ympev.2010.10.004. Accessed 2018-12-12

[2] Zimin, A.V., Puiu, D., Luo, M.-C., Zhu, T., Koren, S., Marçais, G., Yorke, J.A., Dvořák, J., Salzberg, S.L.: Hybrid assembly of the large and highly repetitive genome of Aegilops tauschii, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. Genome Research **27**(5), 787–792 (2017). https://doi.org/10.1101/gr.213405.116

[3] Tremble, K., Suz, L.M., Dentinger, B.T.M.: Lost in translation: Population genomics and long-read sequencing reveals relaxation of concerted evolution of the ribosomal DNA cistron. Molecular Phylogenetics and Evolution **148**, 106804 (2020). https://doi.org/10.1016/j.ympev.2020.106804. Accessed 2021-01-26

[4] Chen, S., Zhou, Y., Chen, Y., Gu, J.: fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics **34**(17), 884–890 (2018). https://doi.org/10.1093/bioinformatics/bty560. Accessed 2019-08-05

[5] Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A.: SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. Journal of Computational Biology **19**(5), 455–477 (2012). https://doi.org/10.1089/cmb.2012.0021. Publisher: Mary Ann Liebert, Inc., publishers. Accessed 2021-09-07

[6] Kohler, A., Kuo, A., Nagy, L.G., Morin, E., Barry, K.W., Buscot, F., Canbäck, B., Choi, C., Cichocki, N., Clum, A., Colpaert, J., Copeland, A., Costa, M.D., Doré, J., Floudas, D., Gay, G., Girlanda, M., Henrissat, B., Herrmann, S., Hess, J., Högberg, N., Johansson, T., Khouja, H.-R., LaButti, K., Lahrmann, U., Levasseur, A., Lindquist, E.A., Lipzen, A., Marmeisse, R., Martino, E., Murat, C., Ngan, C.Y., Nehls, U., Plett, J.M., Pringle, A., Ohm, R.A., Perotto, S., Peter, M., Riley, R., Rineau, F., Ruytinx, J., Salamov, A., Shah, F., Sun, H., Tarkka, M., Tritt, A., Veneault-Fourrey, C., Zuccaro, A., Tunlid, A., Grigoriev, I.V., Hibbett, D.S., Martin, F.: Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. Nature Genetics **47**(4), 410–415 (2015). https://doi.org/10.1038/ng.3223. Bandiera_abtest: a Cc_license_type: cc_y Cg_type: Nature Research Journals Number: 4 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Microbiology Subject_term_id: microbiology. Accessed 2021-09-07

[7] Grigoriev, I.V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F., Smirnova, T., Nordberg, H., Dubchak, I., Shabalov, I.: MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Research **42**(D1), 699–704 (2014). https://doi.org/10.1093/nar/gkt1183. Accessed 2020-04-06

[8] Slater, G.S.C., Birney, E.: Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics **6**(1), 31 (2005). https://doi.org/10.1186/1471-2105-6-31. Accessed 2021-09-07

[9] Katoh, K., Rozewicki, J., Yamada, K.D.: MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Briefings in Bioinformatics (2017). https://doi.org/10.1093/bib/bbx108. Accessed 2019-09-18

[10] Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams,

M.D., von Haeseler, A., Lanfear, R.: IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Molecular Biology and Evolution **37**(5), 1530–1534 (2020). https://doi.org/10.1093/molbev/msaa015. Accessed 2021-09-07

[11] Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S.: ModelFinder: fast model selection for accurate phylogenetic estimates. Nature Methods **14**(6), 587–589 (2017). https://doi.org/10.1038/nmeth.4285. Accessed 2020-04-06

[12] Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., Vinh, L.S.: UFBoot2: Improving the Ultrafast Bootstrap Approximation. Molecular Biology and Evolution **35**(2), 518–522 (2018). https://doi.org/10.1093/molbev/msx281. Accessed 2021-09-07

[13] Zhang, C., Rabiee, M., Sayyari, E., Mirarab, S.: ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics **19**(S6) (2018). https://doi.org/10.1186/s12859-018-2129-y. Accessed 2020-04-06

[14] Ogilvie, H.A., Bouckaert, R.R., Drummond, A.J.: StarBEAST2 Brings Faster Species Tree Inference and Accurate Estimates of Substitution Rates. Molecular Biology and Evolution **34**(8), 2101–2114 (2017). https://doi.org/10.1093/molbev/msx126. Accessed 2021-09-07

[15] He, M.-Q., Zhao, R.-L., Hyde, K.D., Begerow, D., Kemler, M., Yurkov, A., McKenzie, E.H.C., Raspé, O., Kakishima, M., Sánchez-Ramírez, S., Vellinga, E.C., Halling, R., Papp, V., Zmitrovich, I.V., Buyck, B., Ertz, D., Wijayawardene, N.N., Cui, B.-K., Schoutteten, N., Liu, X.-Z., Li, T.-H., Yao, Y.-J., Zhu, X.-Y., Liu, A.-Q., Li, G.-J., Zhang, M.-Z., Ling, Z.-L., Cao, B., Antonín, V., Boekhout, T., da Silva, B.D.B., De Crop, E., Decock, C., Dima, B., Dutta, A.K., Fell, J.W., Geml, J., Ghobad-Nejhad, M., Giachini, A.J., Gibertoni, T.B., Gorjón, S.P., Haelewaters, D., He, S.-H., Hodkinson, B.P., Horak, E., Hoshino, T., Justo, A., Lim, Y.W., Menolli, N., Mešić, A., Moncalvo, J.-M., Mueller, G.M., Nagy, L.G., Nilsson, R.H., Noordeloos, M., Nuytinck, J., Orihara, T., Ratchadawan, C., Rajchenberg, M., Silva-Filho, A.G.S., Sulzbacher, M.A., Tkalčec, Z., Valenzuela, R., Verbeken, A., Vizzini, A., Wartchow, F., Wei, T.-Z., Weiß, M., Zhao, C.-L., Kirk, P.M.: Notes, outline and divergence times of Basidiomycota. Fungal Diversity **99**(1), 105–367 (2019). https://doi.org/10.1007/s13225-019-00435-4. Accessed 2021-09-07

[16] Chen, J.-J., Cui, B.-K., Zhou, L.-W., Korhonen, K., Dai, Y.-C.: Phylogeny, divergence time estimation, and biogeography of the genus Heterobasidion (Basidiomycota, Russulales). Fungal Diversity **71**(1), 185–200 (2015). https://doi.org/10.1007/s13225-014-0317-2. Accessed 2021-09-07

[17] Langmead, B., Salzberg, S.L.: Fast gapped-read alignment with Bowtie 2. Nature Methods **9**(4), 357–359 (2012). https://doi.org/10.1038/nmeth.1923

[18] Kryvokhyzha, D.: GATK: the best practice for genotype calling in a non-model organism. https://evodify.com/gatk-in-non-model-organism/ Accessed 2021-04-07

[19] Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., 1000 Genomes Project Analysis Group: The variant call format and VCFtools. Bioinformatics **27**(15), 2156–2158 (2011). https://doi.org/10.1093/bioinformatics/btr330. Accessed 2021-09-07

[20] Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., Sham, P.C.: PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics **81**(3), 559–575 (2007). https://doi.org/10.1086/519795. Accessed 2021-09-07

[21] Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., François, O.: Fast and Efficient Estimation of Individual Ancestry Coefficients. Genetics **196**(4), 973–983 (2014). https://doi.org/10.1534/genetics.113.160572. Accessed 2021-05-04

[22] Korunes, K.L., Samuk, K.: pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. Molecular Ecology Resources **21**(4), 1359–1368 (2021). https://doi.org/10.1111/1755-0998.13326. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1755-0998.13326. Accessed 2021-07-16

[23] Nelson, C.W., Moncla, L.H., Hughes, A.L.: SNPGenie: estimating evolutionary parameters to detect natural selection using pooled next-generation sequencing data. Bioinformatics **31**(22), 3709–3711 (2015). https://doi.org/10.1093/bioinformatics/btv449. Accessed 2021-07-16

[24] Malinsky, M., Matschiner, M., Svardal, H.: Dsuite - Fast D-statistics and related admixture evidence from VCF files. Molecular Ecology Resources **21**(2), 584–595 (2021). https://doi.org/10.1111/1755-0998.13265. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1755-0998.13265. Accessed 2021-05-17

[25] Luu, K., Bazin, E., Blum, M.G.B.: *pcadapt* : an R package to perform genome scans for selection based on principal component analysis. Molecular Ecology Resources **17**(1), 67–77 (2017). https://doi.org/10.1111/1755-0998.12592. Accessed 2020-03-06

[26] Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E., Yates, C.J.: A statistical explanation of MaxEnt for ecologists. Diversity and Distributions **17**(1), 43–57 (2011). https://doi.org/10.1111/j.1472-4642.2010.00725.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1472-4642.2010.00725.x. Accessed 2021-09-07

[27] Excoffier, L., Marchi, N., Marques, D.A., Matthey-Doret, R., Gouy, A., Sousa, V.C.: *fastsimcoal2* : demographic inference under complex evolutionary scenarios. Bioinformatics, 468 (2021). https://doi.org/10.1093/bioinformatics/btab468. Accessed 2021-09-07

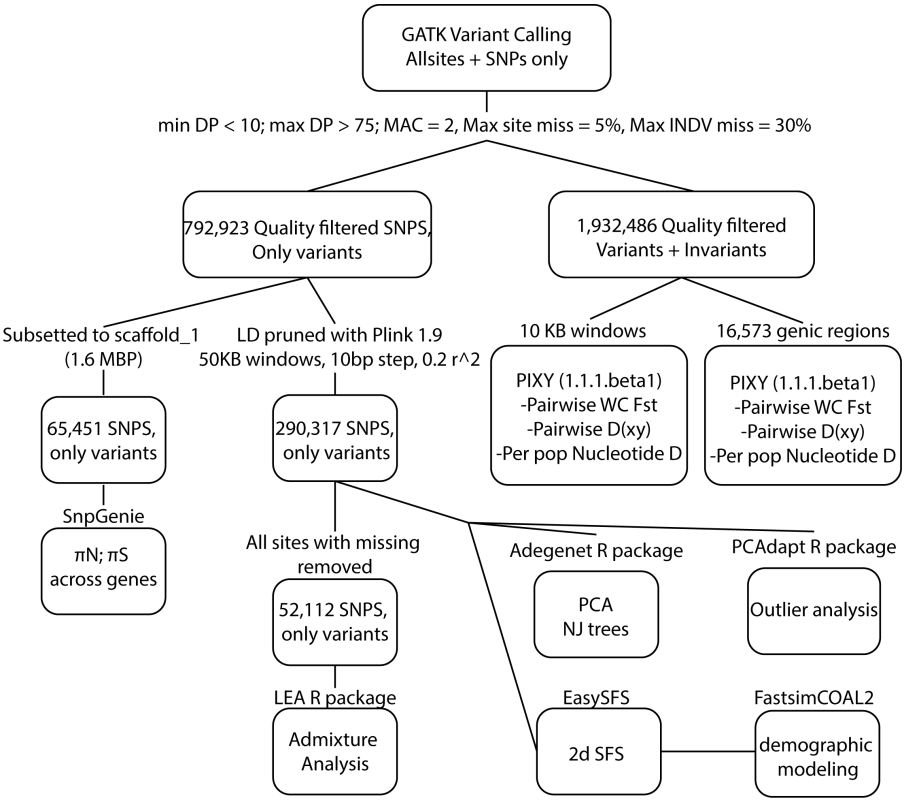

Fig. 1 Flowchart of population genomic methods and SNP datasets used for each analysis.

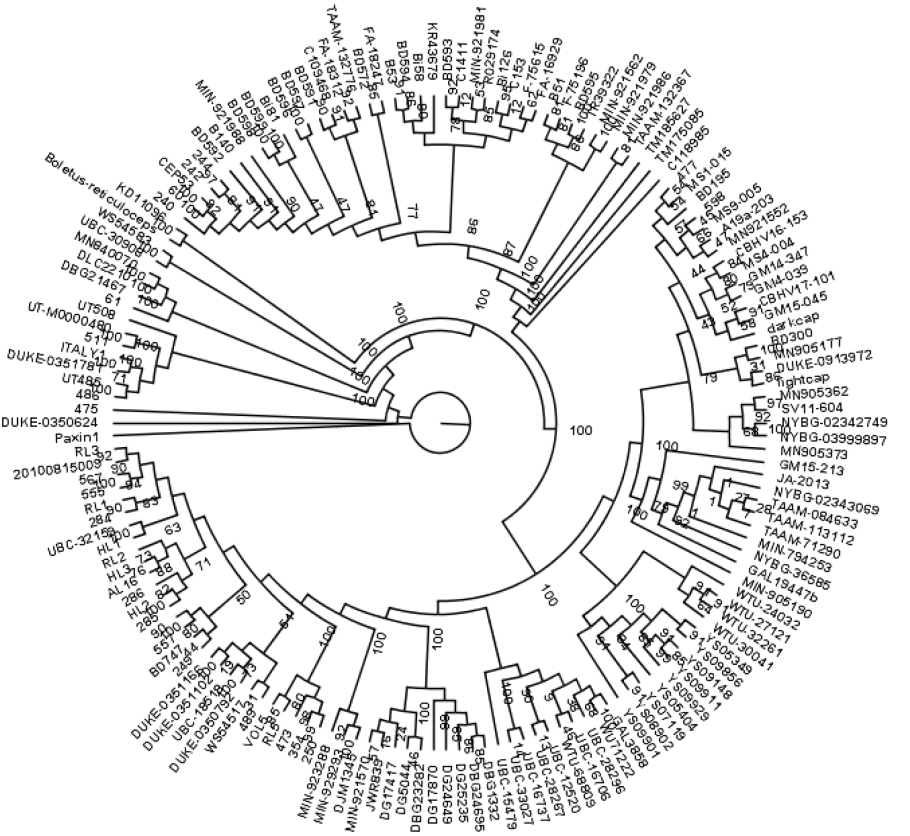

Fig. 2 Concatenated supermatrix phylogenomic analysis using the 702 conserved single copy dataset. Node labels represent bootstrap support.

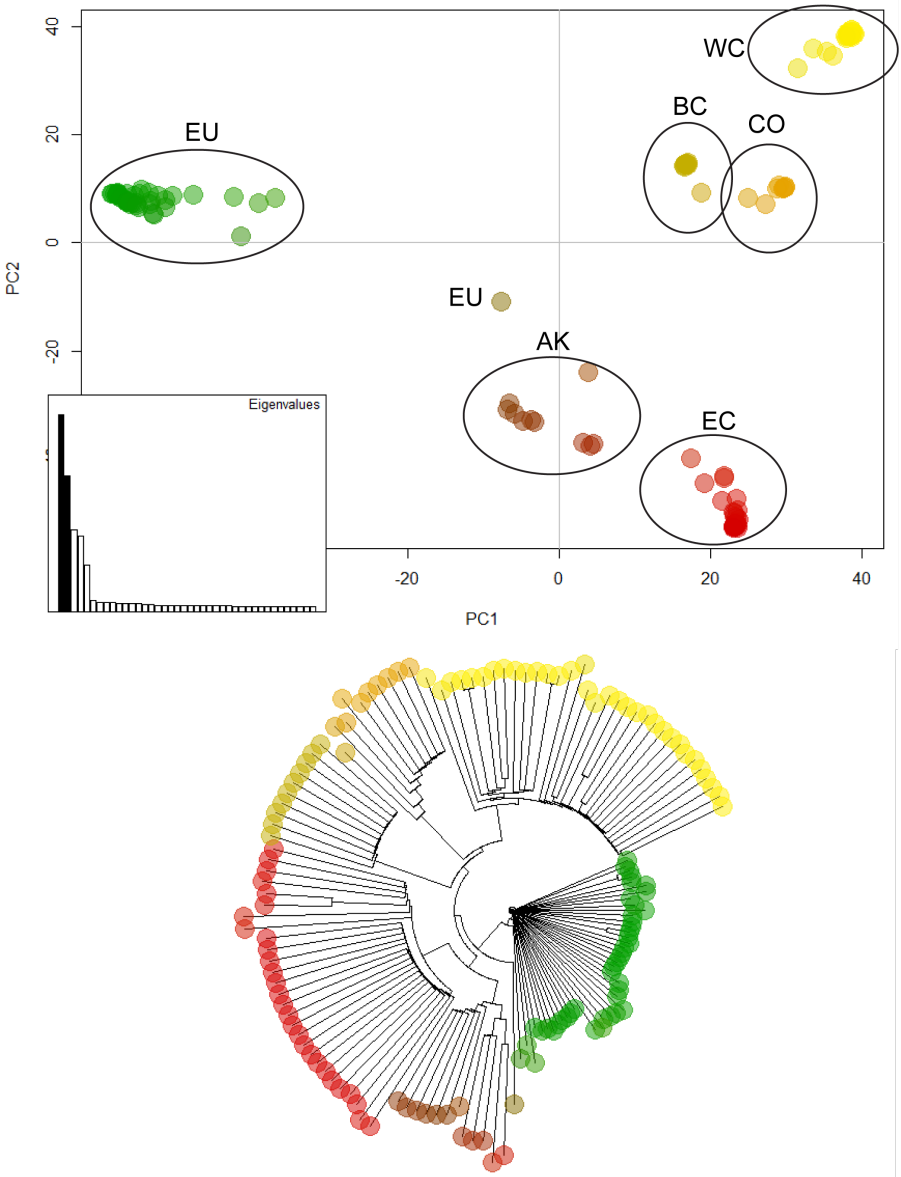

Fig. 3 Population structure of *B. edulis lineages assessed with top) principle component analysis (PCA) using the first two principle components, and bottom) neighbor-joining tree with tips colored based on results from PCA. The six lineages recovered from phylogenomic and admixture analysis were also recovered in these assessments.*

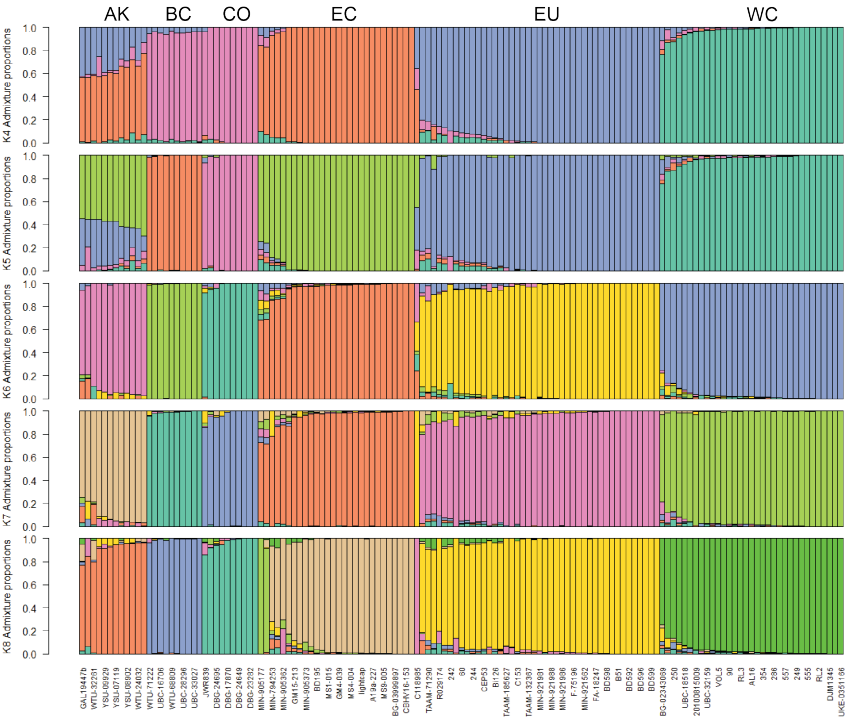

Fig. 4 Stacked admixture analysis performed with the R package LEA for K=4 through K=8. At K=4 and K=5 the Ak lineage is clustered with the EC and EU lineages, indicating recent admixture within these lineages. At high vales of K (K¿6) the 6 lineages are surprisingly still retained, with only one or two individuals being segregated into new clusters. This further indicates little lineage substructure.

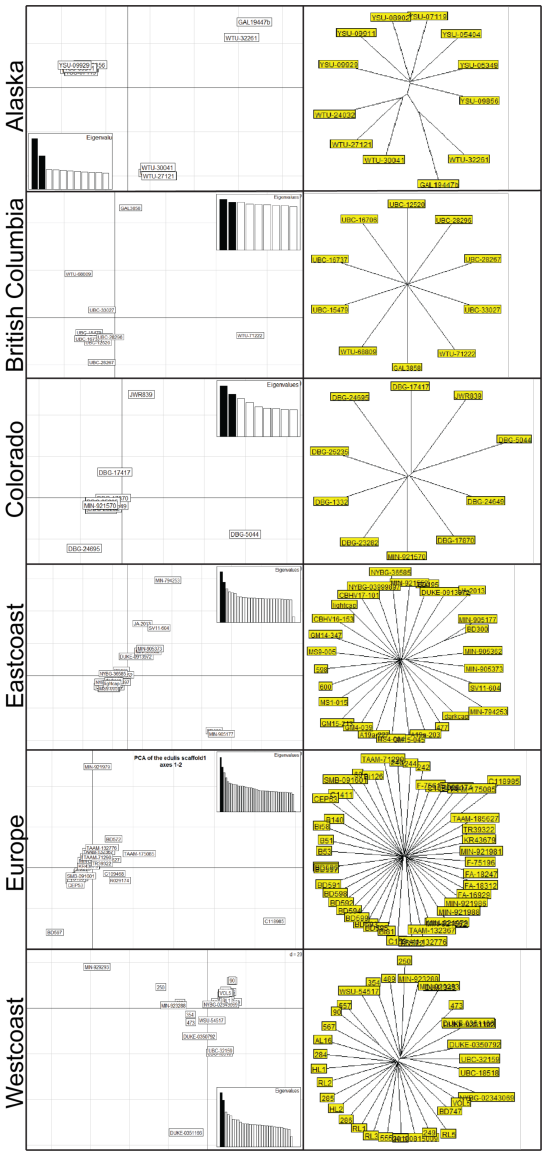

Fig. 5 Principle component analysis and neighbor-joining trees of within lineage population for all 6 lineages identified. Little within-lineage substructure is found outside of the AK lineage.

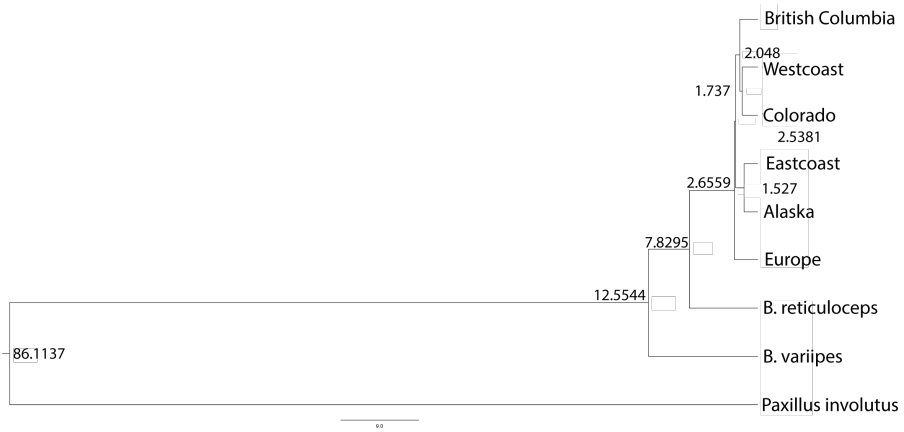

Fig. 6 Divergence date estimation using StarBEAST2 and 20 randomly selected loci from the 702 single copy dataset. All parameters reached ESS ¿ 200 which indicates adequate convergence.

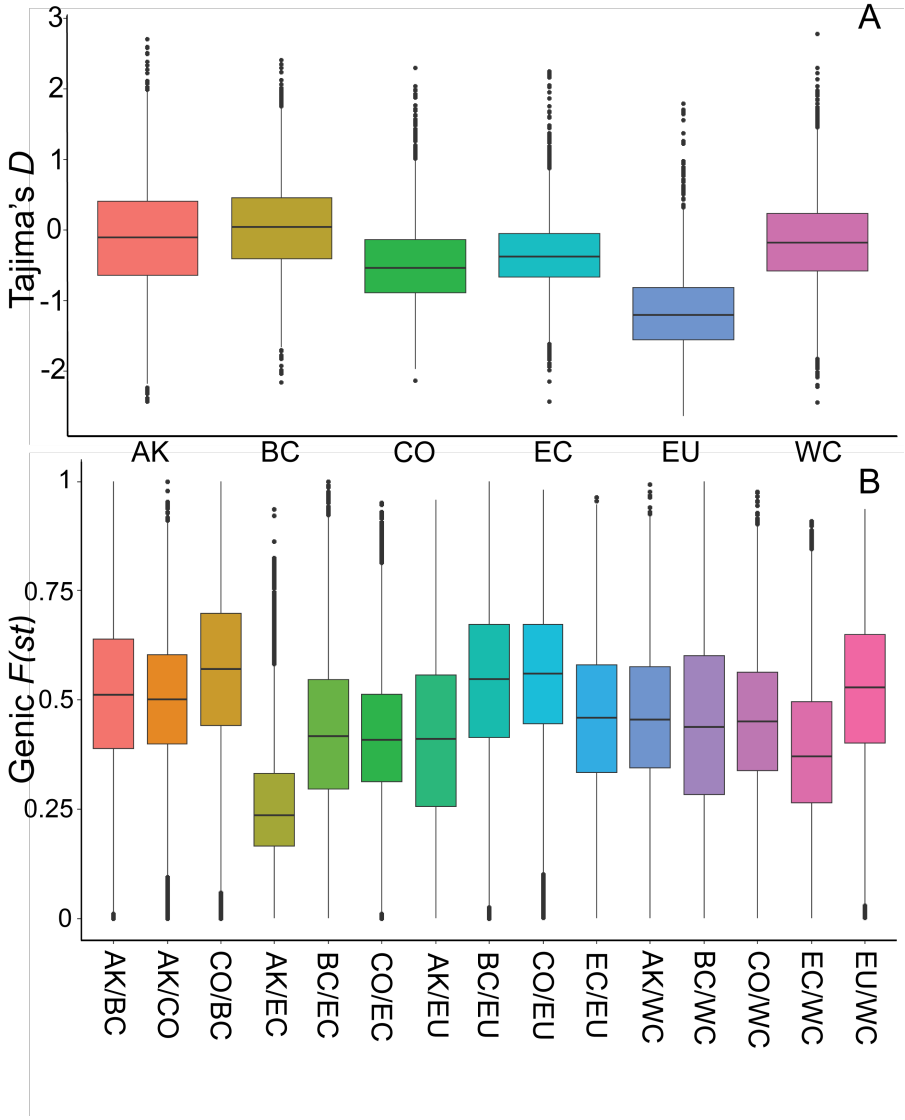

Fig. 7 A) Tajima's D statistic calculated with VCFTOOLS. The EU lineage exhibits a significantly (P¡ 0.05) lower, negative D value, possibly indicating rapid population expansion. B) Pairwise population *F(st)* calculated with PIXY across only genic regions. High *F(st)* values for most comparisons indicate that divergence is genome-wide and not relegated to genomic regions of low selective constraint such as intergenic regions.

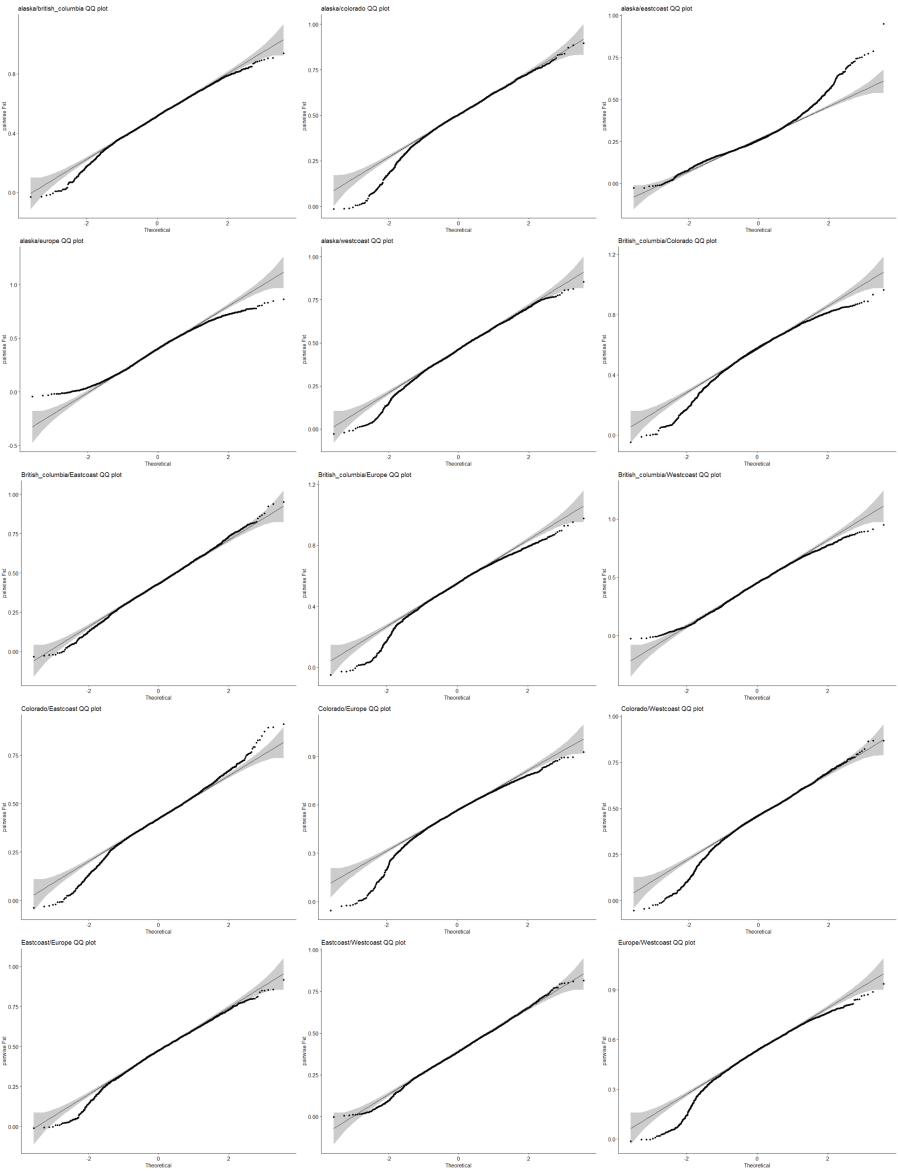

Fig. 8 QQplot distributions of 10kb window pairwise *F(st)* values. Ongoing gene-flow can reduce genome-wide differentiation around all but select "divergence islands" under divergent-selection. The cessation of gene-flow relaxes genomic homogenization, allowing divergence to accumulate in all but the most-conserved loci. All pairwise comparisons except AK/EC have distributions that can indicate minimal ongoing geneflow, increased abundance than expected of loci with little divergence due to high conservation. AK/EC have an increased abundance than expected of highly diverged loci, indicating genomic divergence islands.

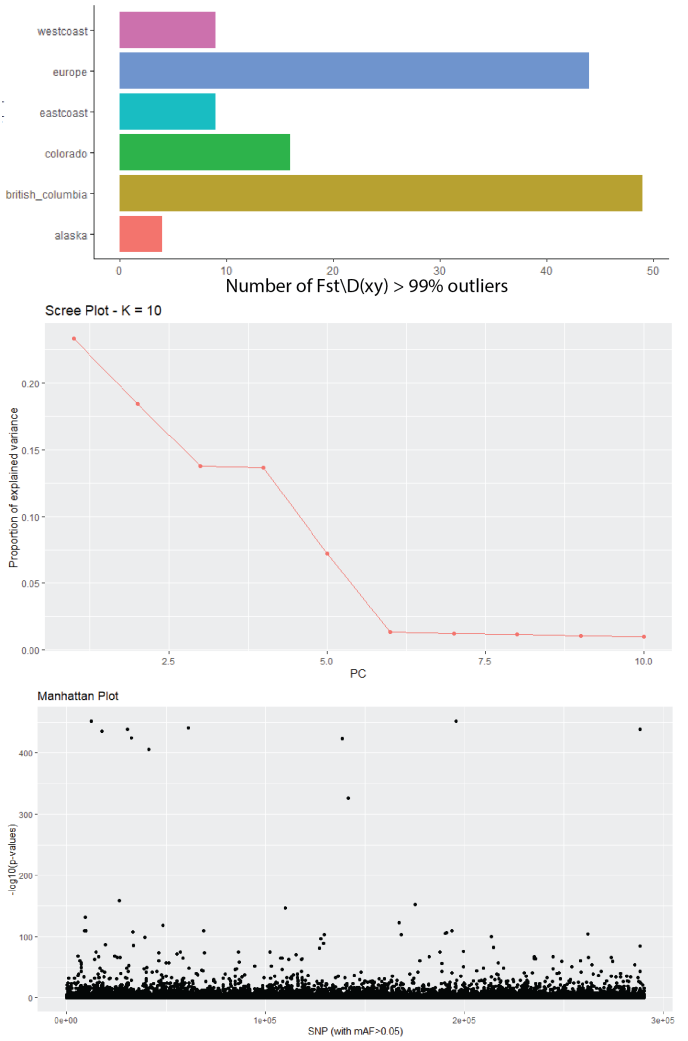

Fig. 9 Results of genome-wide outlier analysis to identify divergence loci. Top) The number of highly divergence loci, in top 1% of divergence of all loci, found within each population. Middle) Scree plot produced by PCAdapt indicating the proportion of variance explained by K principle components. K=3 was selected according to proscribed protocol as this represented the first "knee" plateau, potentially representing the three branches within the *B. edulis* phylogeny. Bottom) Manhatttan plot produced by PCAdapt indicating the significance of putative locally-adapted loci.

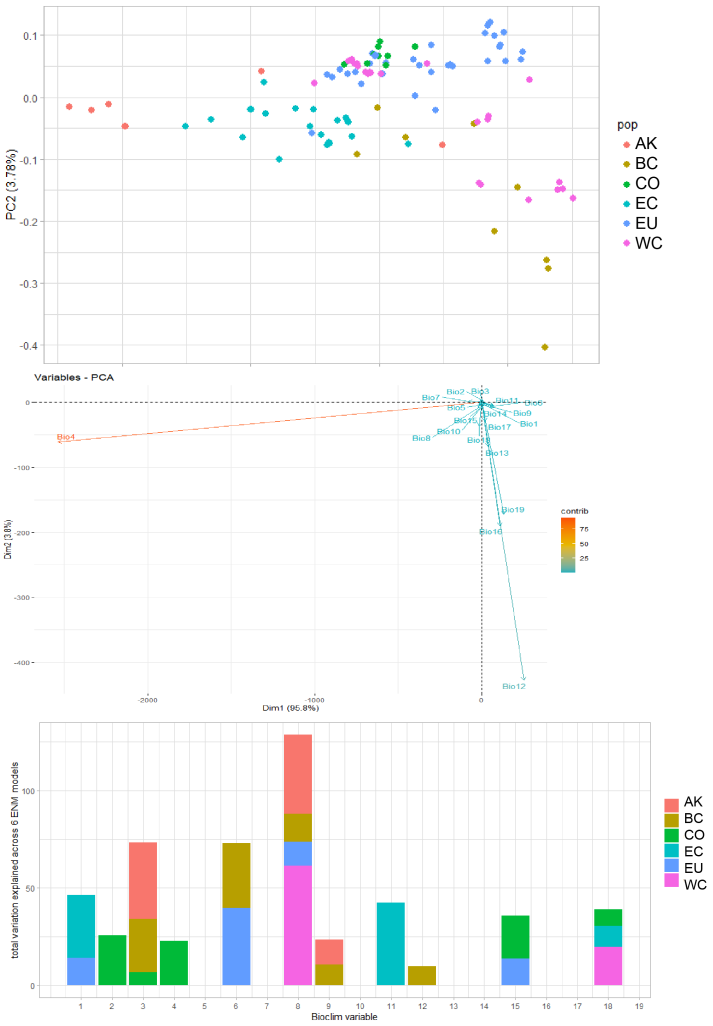

Fig. 10 Environmental coorelation analysis of B. edulis lineages using A) 19 Bioclim values extracted for the location of each specimen and assessed using PCA, B) Biplot of 19 Bioclim variables indicating that Bio4 is primary variable contributing to variation between specimens, C) Contribution of Bioclim variables to each lineage Maxent ecological niche model.

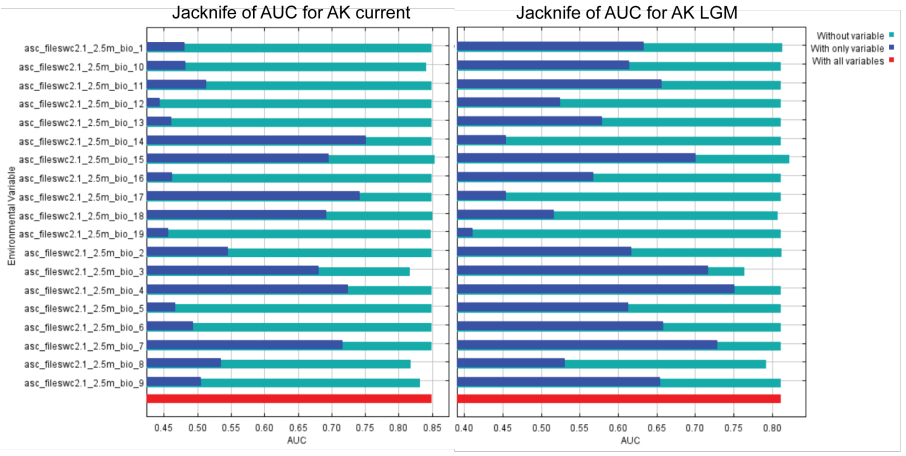

Fig. 11 Jacknife contributions of each Bioclim variable to the AUC of the AK ecological niche model for left) current conditions and right) conditions during the last glacial maximum (22Kya).

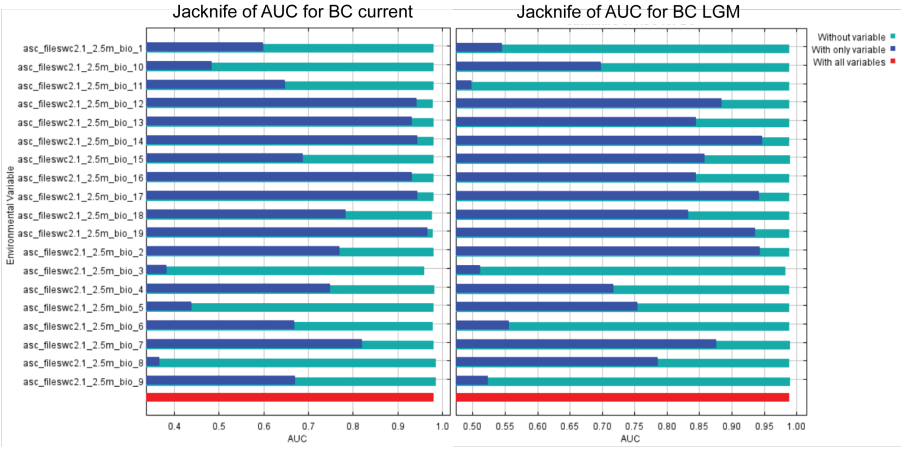

Fig. 12 Jacknife contributions of each Bioclim variable to the AUC of the BC ecological niche model for left) current conditions and right) conditions during the last glacial maximum (22Kya)

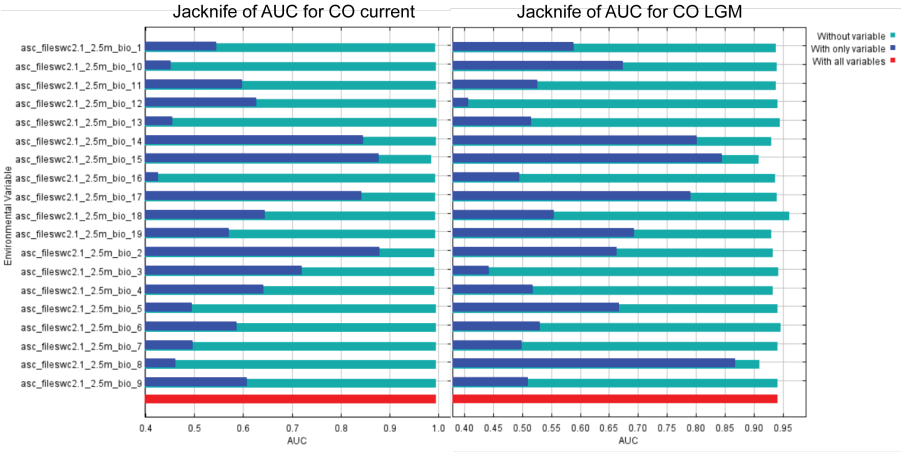

Fig. 13 Jacknife contributions of each Bioclim variable to the AUC of the CO ecological niche model for left) current conditions and right) conditions during the last glacial maximum (22Kya)

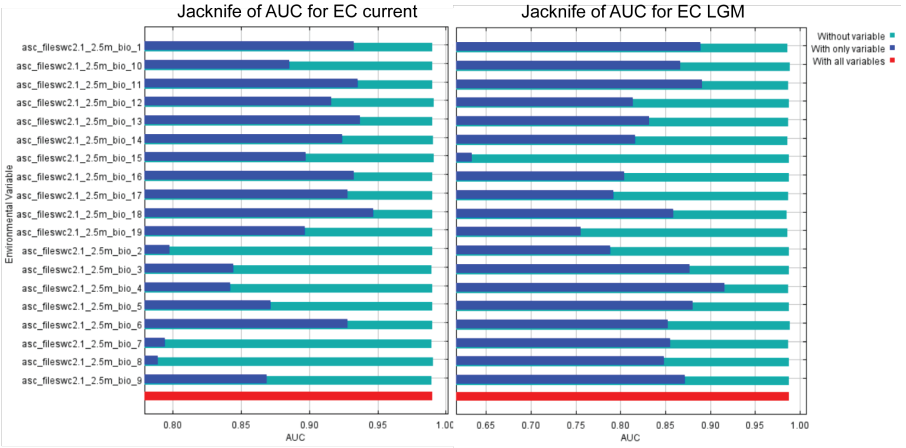

Fig. 14 Jacknife contributions of each Bioclim variable to the AUC of the EC ecological niche model for left) current conditions and right) conditions during the last glacial maximum (22Kya)

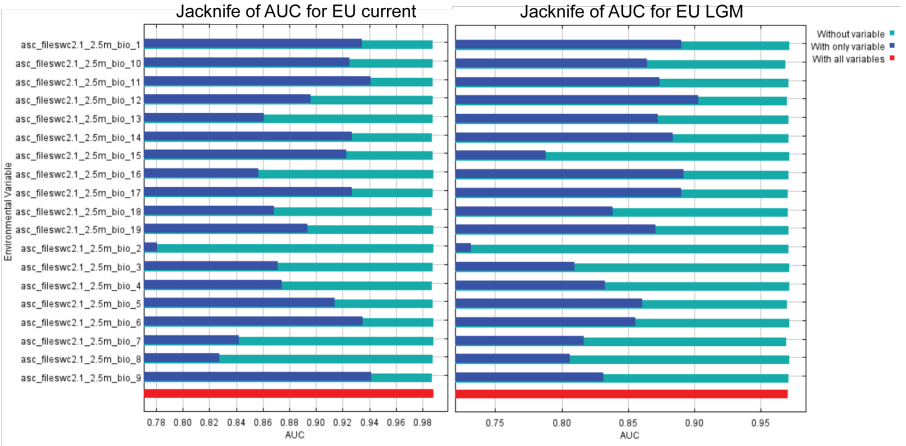

Fig. 15 Jacknife contributions of each Bioclim variable to the AUC of the EU ecological niche model for left) current conditions and right) conditions during the last glacial maximum (22Kya)

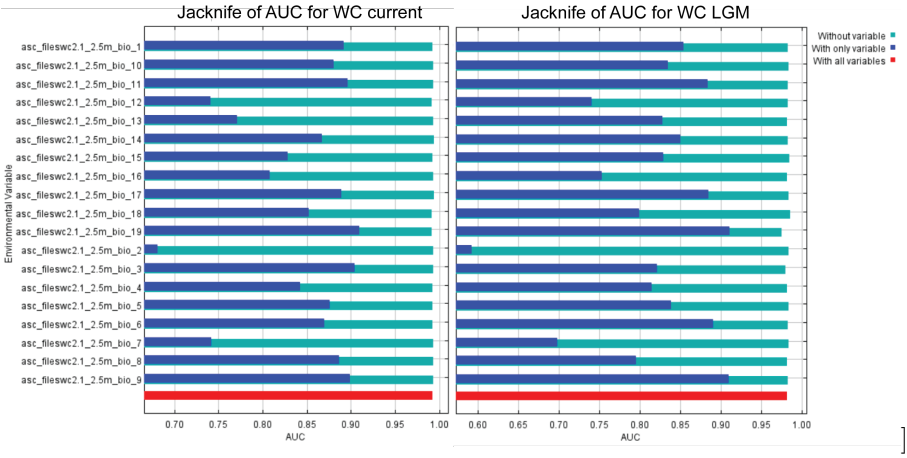

Fig. 16 Jacknife contributions of each Bioclim variable to the AUC of the EC ecological niche model for left) current conditions and right) conditions during the last glacial maximum (22Kya)