**Supplementary Material**

**Table S1**. **Metadata for 11 Atlantic salmon samples sequenced with nanopore long-read technology.** Wild salmonwere sampled to represent the four main phylogeographic groups; North American (NAm), Baltic (BAL), Barents/White Sea (BWS) and Atlantic (ATL). The aquaculture sample (AQGE) was sampled from the AquaGen strain originating mainly from the ATL group.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Name** | **Species** | **River name** | **Phylo. group** | **Country** | **Gender** | **Pop. type** | **Lat,**  **Long** | **ENA Project**  **accession** |
| AQGE | Altantic salmon | - | - | Norway | Male | Aquaculture | - | PRJEB43080 |
| GLOP | Altantic salmon | Gloppenelva | ATL | Norway | Male | Anadromous | 61.46N, 6.12E | PRJEB50984 |
| ARUN | Altantic salmon | Årungselva | ATL | Norway | Male | Anadromous | 59.43N, 10.43E | PRJEB50985 |
| ALTA | Altantic salmon | Altaelva | BWS | Norway | Male | Anadromous | 69.58N, 23.22E | PRJEB50986 |
| TANA | Altantic salmon | Tanaelva | BWS | Norway | Male | Anadromous | 70.29N, 28.23E | PRJEB50987 |
| FROM | Altantic salmon | River Frome | ATL | UK | Male | Anadromous | 50.41N, 2.05W | PRJEB50988 |
| OULO | Altantic salmon | Oulujoki | BAL | Finland | Male | Anadromous | 64.98N, 25.61E | PRJEB50989 |
| PERU | Altantic salmon | Lac Perugia | NAm | Canada | Male | Landlocked | 47.43N, 76.30W | PRJEB50990 |
| SEBA | Altantic salmon | Sebago Lake | NAm | USA | Female | Landlocked | 43.52N, 70.34W | PRJEB50991 |
| GARN-1 | Altantic salmon | Garnish River | NAm | Canada | Male | Anadromous | 47.23N, 55.35W | PRJEB49548 |
| GARN-2 | Altantic salmon | Garnish River | NAm | Canada | Male | Anadromous | 47.23N, 55.35W | PRJEB50992 |
| ARUN | Brown trout | Årungselva | ATL | Norway | Male | Anadromous | 59.43N, 10.43E | PRJEB50994 |

**Table S2. Indel enrichment in inversion regions.** The enrichment was considered as differences between inversion region and homologous region in the genome divided by length of the regions counted and as; i) number of indels and ii) base pairs in indels.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | **North America** | | **Europe** | |
| **Inversion** | **Homeologous region** | **i) Δ number of indels** | **ii) Δ base pairs in indels** | **i) Δ number of indels** | **ii) Δ base pairs in indels** |
| chr3inv | chr5:21,053,100-21,112,740 | 2.0 | 1.3 | 4.1 | 1.2 |
| chr4inv | chr13:76,443,570-76,538,610 | 1.0 | 1.4 | 0.0 | 0.3 |
| chr09inv | chr5:10,657,170-17,305,100 | 1.0 | 1.1 | 1.1 | 1.0 |
| chr10inv | chr16:3,100,000-3,300,000 | 0.4 | 0.8 | 1.6 | 0.7 |
| chr11inv1 | No homeologous region | - | - | - | - |
| chr11inv2 | No homeologous region | - | - | - | - |
| chr11inv3 | chr26:53,135,590-54,987,680 | 0.6 | 0.5 | 0.6 | 0.5 |
| chr16inv | No homeologous region | - | - | - | - |
| chr18inv | chr7:27,565,350-29,960,450 | 3.4 | 5.0 | 2.7 | 2.6 |
| chr22inv | chr12:92,302,090-93,554,530 | 0.7 | 1.0 | 1.2 | 0.6 |
| chr26inv | chr11:53,066,290-56,862,090 | 2.5 | 4.3 | 1.9 | 1.9 |

**Table S3. Environment correlation matrix.** Correlation (r2) among environmental variables in European (above diagonal) and North American (below diagonal) populations. The correlation is high between mean annual temperature and temperature in the coldest quarter in Europe and the warmest quarter in North America. Mean annual temperature and latitude are highly correlated in both groups.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Lat** | **Basin size** | **Temp (mean)** | **Temp (warmQ)** | **Temp (coldQ)** | **IsoTherm** | **Precip** |
| **Lat** |  | -0.14 | -0.9 | -0.83 | -0.75 | -0.6 | -0.59 |
| **Basin size** | 0.41 |  | -0.01 | 0.38 | -0.18 | -0.29 | -0.19 |
| **Temp (mean)** | -0.97 | -0.36 | - | 0.75 | 0.95 | 0.66 | 0.7 |
| **Temp (warmQ)** | -0.94 | -0.32 | 0.93 | - | 0.51 | 0.25 | 0.37 |
| **Temp (coldQ)** | -0.37 | -0.21 | 0.5 | 0.14 | - | 0.72 | 0.74 |
| **IsoTherm** | 0.23 | 0.01 | -0.2 | -0.45 | 0.43 | - | 0.58 |
| **Precip** | 0.61 | 0.17 | -0.45 | -0.44 | -0.07 | -0.13 | - |

Chart, scatter chart

Description automatically generated

**Figure S1: Tandem repeat structures in inversion breakpoints.** Self-alignments of 11 inversion sequences in Atlantic salmon visualizing inversion breakpoints. Diagonal solid lines indicate breakpoint coordinates for **A** (chr3inv), **C** (chr9inv), **F** (chr11inv3), **I** (chr22inv) and **J** (chr26inv) with repeat structures at the breakpoints, and **B** (Chr4inv), **D** (chr10inv), **E** (chr11inv1 and chr11inv2), **G** (chr16inv) and **H** (chr18inv) without obvious tandem repeat structures at inversion breakpoints. **C** (Chr9inv) and **F** (chr11inv3) display 200kbp up- and downstream of the breakpoints with boarder between the sequences marked with a dashed line.

**Chart, scatter chart

Description automatically generated**

**Figure S2. Confirmation of inversion structure of chr9inv using long-reads and contigs spanning the upstream breakpoint.** **A**. Self-alignment of chr9inv in sample AQGE used to generate the Atlantic salmon reference sequence GCA\_905237065.2. **B**. Ultra-long reads spanning the upstream breakpoint chr9inv used to determine that AQGE is heterozygous for the inversion. **C**. Contigs spanning the upstream inversion breakpoint in the OULO sample, validating the alternative state of the inversion in this sample.

**Chart, treemap chart

Description automatically generated**

**Figure S3. Visualization of haplotype structures within inversion regions.** SNP genotypes were clustered by the hierarchical clustering methods per sample. Navy: alternative homozygous, white: heterozygous, and red: reference homozygous SNPs. Individuals used to construct long-read assemblies are highlighted in black in the left bar, these were used to test if the structure reflects inversion orientation. Variants in 5kb up/downstream and the inversion were used (more than 5% minor allele frequency). Colors indicate the different phylogeographical lineages, Atlantic, White/Barents Sea, Baltic and North American. Chr4inv in European populations and chr18inv in North American populations are the only haplotype structures following the inversion pattern. No SNPs could be called for chr16iv.

**Chart

Description automatically generated**

**Figure S4. Comparisons of Atlantic salmon inversions with brown trout to determine ancestral state.** Comparison of inversions in Atlantic salmon reference AQGE with syntenic regions in brown trout. Smaller inversions (<1Mbp) were aligned with LASTZ (**A**, **B**, **D**, **E** and **G**) and larger with Minimap2 (**C**, **F**, **H**, **I** and **J**). Chr3inv (**A**), chr9inv (**C**), chr10inv (**D**) and chr11inv1&2 (**E**) show inverted orientation, while the remaining regions show ancestral orientation of the inversions. Colour coding shows alignment orientation (red – and blue +).

**Diagram

Description automatically generated**

**Figure S5. Gene annotation in chr18inv upstream breakpoint.** Ensembl Rapid Release annotation (Ssal\_v3.1 version 104.1) for the upstream breakpoint of chr18inv. **A.** The chr18inv breakpoint, indicated by vertical red stapled line, disrupt ENSSSAG00000081762 (C-type mannose receptor 2-like) by breaking in intron 1 of the gene. **B.** Two other C-type mannose receptor 2-like genes in the close vicinity of ENSSSAG00000081762, ENSSSAG00000054546 and ENSSSAG00000054554, are not directly affected by the inversion.

A picture containing chart

Description automatically generated

**Figure S6a**. **Gene annotation in chr22inv upstream breakpoint.** Ensembl Rapid Release annotation (Ssal\_v3.1 version 104.1) for the upstream breakpoint of chr22inv, indicated by vertical red stapled line, disrupt the genes ENSSSAG00000067423 annotated as protein-glutamine gamma-glutamyltransferase 2-like isoform X1 (*TGM2-like*) and ENSSSAG00000094735 annotated as serine/threonine-protein kinase (*VRK3*).

**Table

Description automatically generated with medium confidence**

**Figure S6b**. **Gene annotation in chr22inv downstream breakpoint.** Ensembl Rapid Release annotation (Ssal\_v3.1 version 104.1) for the downstream breakpoint of chr22inv, indicated by vertical red stapled line, disrupt the gene ENSSSAG00000090871 annotated as deoxyribonuclease gamma-like isoform X3 (*DNASE1L3*).

A picture containing graphical user interface

Description automatically generated

**Figure S7**. **Gene annotation in chr26inv downstream breakpoint.** Ensembl Rapid Release annotation (Ssal\_v3.1 version 104.1) for the downstream breakpoint of chr26inv, indicated by vertical red stapled line, disrupt the gene ENSSSAG00000107384 annotated as HLA-B associated transcript 1 (*BAT1*).

Chart

Description automatically generated

**Figure S8. Accumulation of deleterious mutations in inversions.** Boxplot of the number of deleterious mutations per megabase [0,3] in inversions containing genes compared to the rest of the genome for **A** variants detected in European populations, and **B** variants detected in North American populations.

Graphical user interface, text, application

Description automatically generated

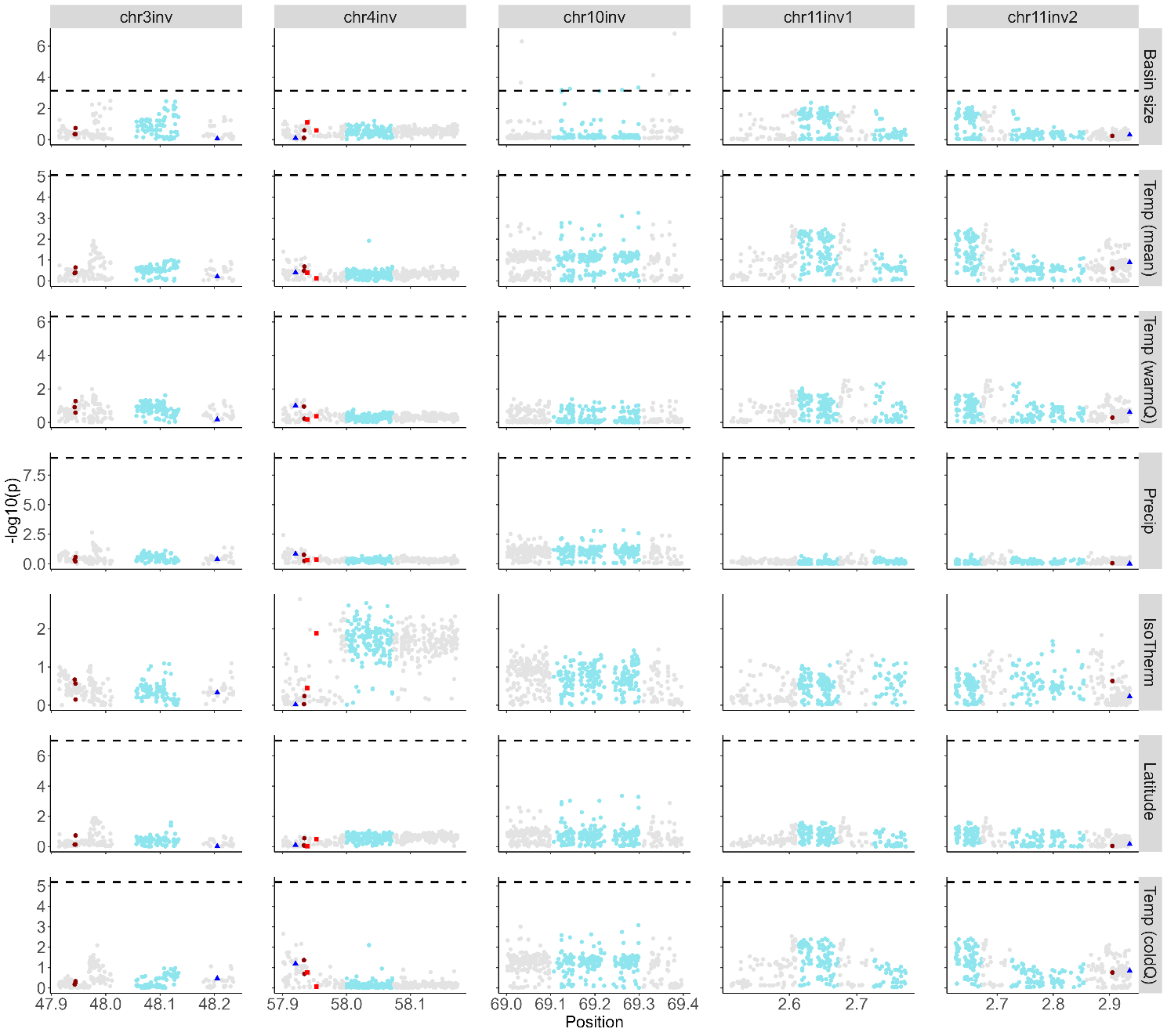
**Figure S9. Deletion segregating with chr18inv.** Location of a 260 bp deletion within chr18inv overlapping the 3’-end of ENSSSAG00000044266 annotated as *P2RY5* (P2Y purinoceptor 5) in the Ensembl Rapid Release annotation (Ssal\_v3.1 version 104.1). The deletion, segregating perfectly with the ancestral configuration of chr18inv in North America populations, is indicated by a red stapled box in the figure.



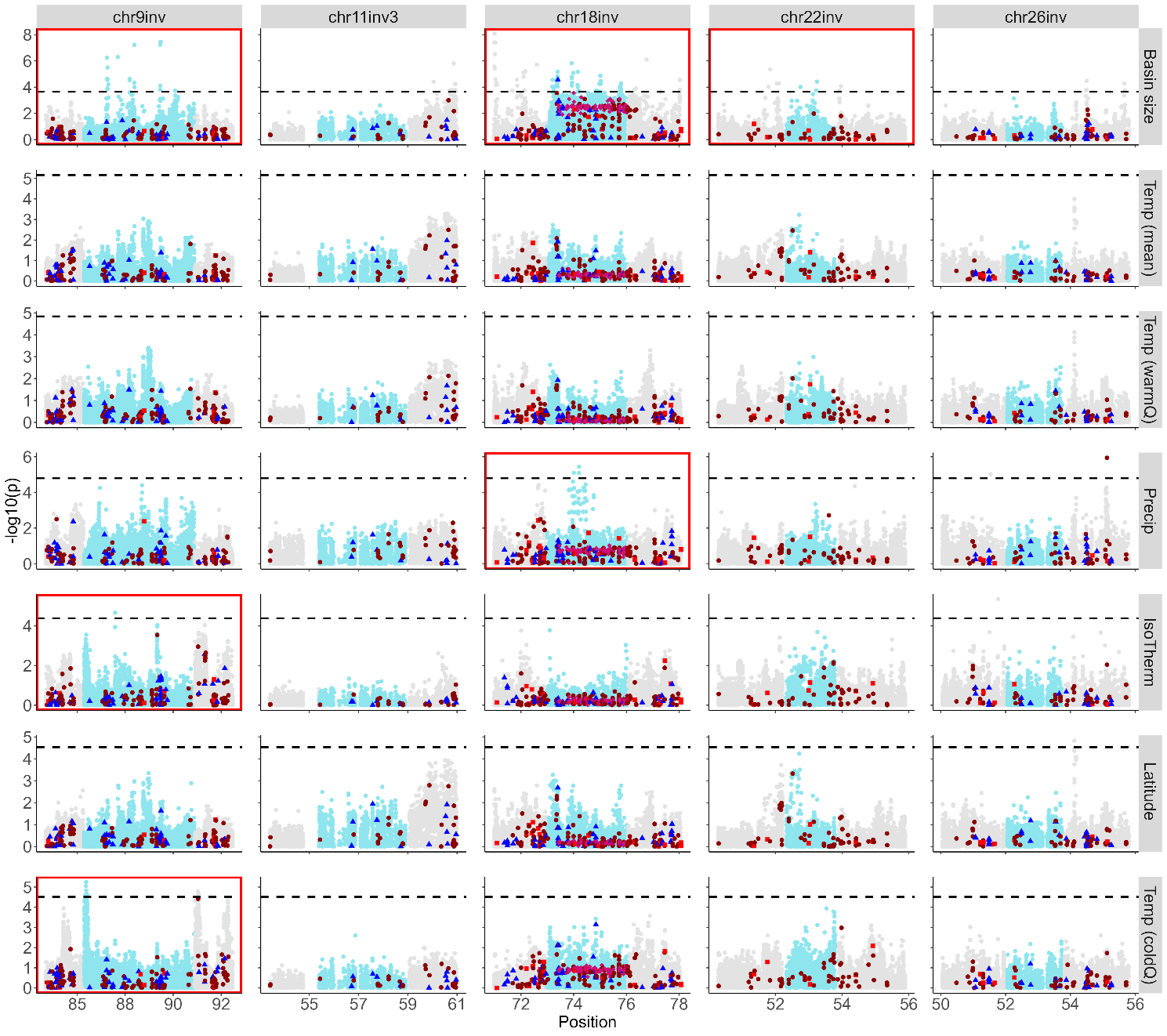
**Figure S10. Dating of chr18inv.** Split plot (smc++) showing estimated date of origin of the inversion. The red line represents the ancestral homozygotes and the blue line the inverted homozygotes. The inversion is estimated to have originated ~5000 generations ago, which is equivalent to approximately 15,000 years with a three-year generation time. This date is when the glacial retreat started, and Atlantic salmon began a postglacial range expansion.



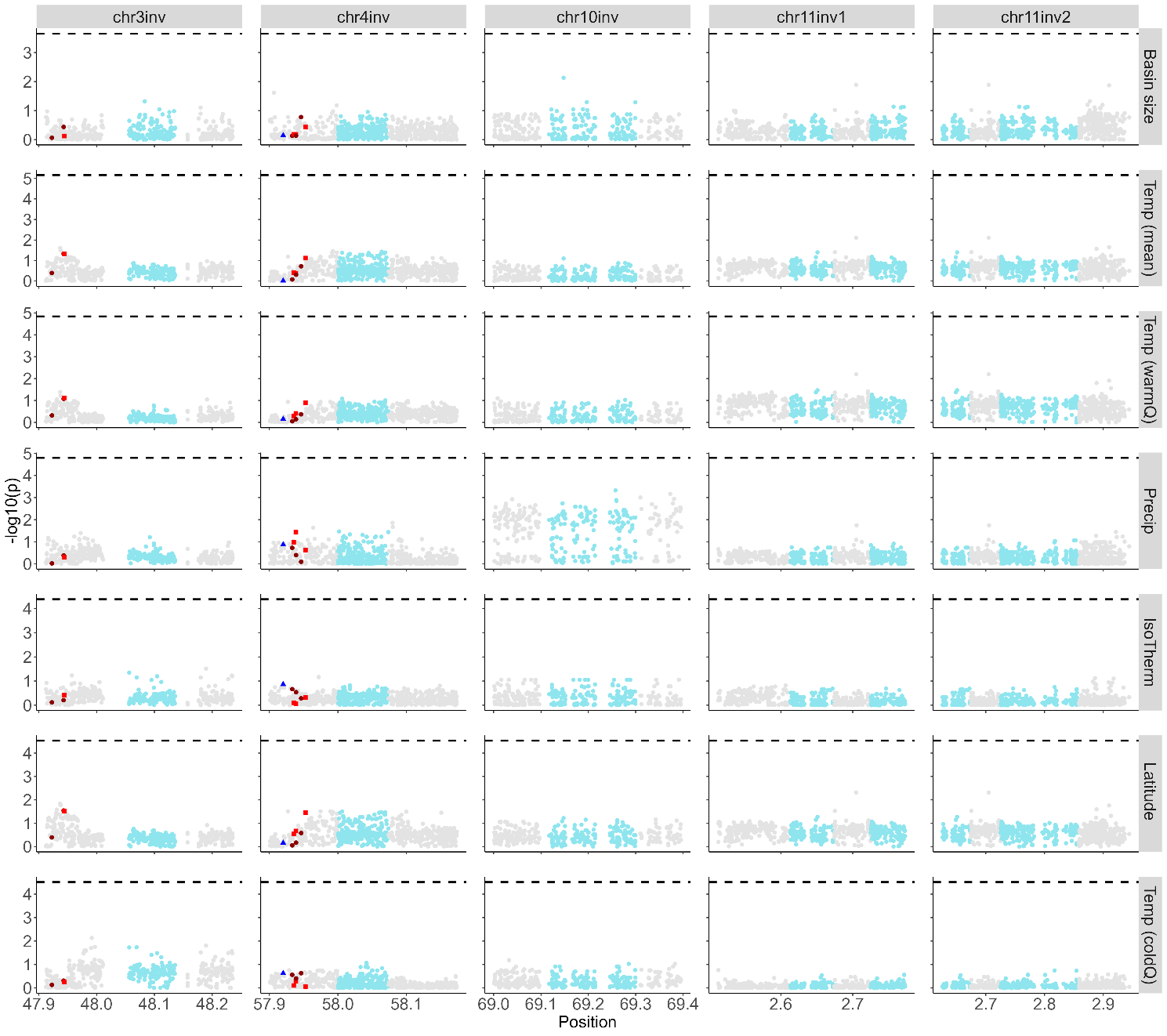
**Figure S11a. Genotype-environment association for SNPs found in European populations within larger inversions (>1Mb)**. SNPs inside inverted sequence is blue and 2Mb and up- and downstream flanking sequence is grey for; drainage basin area (sqkm), annual mean temperature, mean temperature of warmest quarter, annual precipitation, isothermality, latitude and mean temperature of coldest quarter. Dashed horizontal line marks the adjusted p < 0.05 significance threshold. Subplots with significant SNPs are marked with red frames. A significance threshold could not be calculated for annual precipitation. Red squares show missense variants annotated as “high” impact by SNPeff, dark red circles show “moderate” impact, and blue triangles mark deleterious mutations (≥|2.5| PROVEAN scores).

****

**Figure S11b. Genotype-environment association for SNPs found in European populations within smaller inversions (<1Mb).** SNPs inside inverted sequence is blue and 2Mb and up- and downstream flanking sequence is grey for; drainage basin area (sqkm), annual mean temperature, mean temperature of warmest quarter, annual precipitation, isothermality, latitude and mean temperature of coldest quarter. Dashed horizontal line marks the adjusted p < 0.05 significance threshold. Subplots with significant SNPs are marked with red frames. A significance threshold could not be calculated for annual precipitation. Red squares show missense variants annotated as “high” impact by SNPeff, dark red circles show “moderate” impact, and blue triangles mark deleterious mutations (≥|2.5| PROVEAN scores).

****

**Figure S11c. Genotype-environment association for SNPs found in North American populations within larger inversions (>1Mb).** SNPs inside inverted sequence is blue and 2Mb and up- and downstream flanking sequence is grey for; drainage basin area (sqkm), annual mean temperature, mean temperature of warmest quarter, annual precipitation, isothermality, latitude and mean temperature of coldest quarter. Dashed horizontal line marks the adjusted p < 0.05 significance threshold. Subplots with significant SNPs are marked with red frames. Red squares show missense variants annotated as “high” impact by SNPeff, dark red circles show “moderate” impact, and blue triangles mark deleterious mutations (≥|2.5| PROVEAN scores). Pink diamonds represent TAG-SNPs for chr18inv.

****

**Figure S11d. Genotype-environment association for SNPs found in North American populations within smaller inversions (<1Mb)**. SNPs inside inverted sequence is blue and 2Mb and up- and downstream flanking sequence is grey for; drainage basin area (sqkm), annual mean temperature, mean temperature of warmest quarter, annual precipitation, isothermality, latitude and mean temperature of coldest quarter. Dashed horizontal line marks the adjusted p < 0.05 significance threshold. Subplots with significant SNPs are marked with red frames. Red squares show missense variants annotated as “high” impact by SNPeff, dark red circles show “moderate” impact, and blue triangles mark deleterious mutations (≥|2.5| PROVEAN scores).

# Supplementary methods

*Long-read based SV-detection*

We detected inversions in the long-read sequenced samples with both read-mapping and assembly comparisons. Read mapping was performed with Winnowmap2 [1], using AQGE as a reference. Sam-files were sorted and converted into BAM-files with Samtools v1.3.1 [2]. The calling was carried out with three separate long-read SV-calling programs, Sniffles v1.2.12 [3], SVIM v1.2.0 [4] and NanoVar 1.3.9 [5], using default settings for SVIM and NanoVar. The minimum number of reads required (-s) was set to 1/3 of the median length when running Sniffles. SVs called as type ‘breakpoint’, i.e. unresolved variants, and other excess information was filtered out using custom scripts available at <https://github.com/kristinastenlokk/long_read_SV>. To increase accuracy, VCFs were merged across program with Jasmine v1.1.0 [6], retaining only variants detected with at least two programs.

Inversion calls were additionally filtered by detection in at least two samples and a lower size limit was set to 10kb. Duplicated and overlapping variants were filtered out by stringent manual curation.

[1] Jain, C., Rhie, A., Hansen, N., Koren, S. & Phillippy, A. M. J. b. 2020 A long read mapping method for highly repetitive reference sequences. *bioRxiv*. (DOI:<https://doi.org/10.1101/2020.11.01.363887>).

[2] Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & Durbin, R. J. B. 2009 The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078-2079. (DOI:<https://doi.org/10.1093/bioinformatics/btp352>).

[3] Sedlazeck, F. J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., Von Haeseler, A. & Schatz, M. C. J. N. m. 2018 Accurate detection of complex structural variations using single-molecule sequencing. *Nature methods* **15**, 461-468. (DOI:<https://doi.org/10.1038/s41592-018-0001-7>).

[4] Heller, D. & Vingron, M. J. B. 2019 SVIM: structural variant identification using mapped long reads. *Bioinformatics* **35**, 2907-2915. (DOI:<https://doi.org/10.1093/bioinformatics/btz041>).

[5] Tham, C. Y., Tirado-Magallanes, R., Goh, Y., Fullwood, M. J., Koh, B. T., Wang, W., Ng, C. H., Chng, W. J., Thiery, A. & Tenen, D. G. J. G. b. 2020 NanoVar: accurate characterization of patients’ genomic structural variants using low-depth nanopore sequencing. *Genome biology* **21**, 1-15. (DOI:<https://doi.org/10.1186/s13059-020-01968-7>).

[6] Kirsche, M., Prabhu, G., Sherman, R., Ni, B., Aganezov, S. & Schatz, M. C. J. B. 2021 Jasmine: Population-scale structural variant comparison and analysis. *bioRxiv*. (DOI:<https://doi.org/10.1101/2021.05.27.445886>).