FIG. S1. Source of experimental animals. (**A**) Broodstock were obtained from two freshwater populations: Lake Vättern (VAT) in central Sweden and Lake Pulmanki (Pulmankijärvi, PUL) in Finnish Lapland. A single marine population was also sampled from the Baltic Sea in the vicinity of Helsinki (HEL), Finland. Map grids, both latitude and longitude, are in 5° intervals. (**B**) Schematic representation of the half-sib breeding design used in quantitative genetic analyses. In this design a single male was mated to two unique females. In total, 30 such half-sib blocks were produced, resulting in 60 families. Individuals sampled randomly from each family were subjected to one of two temperature treatments immediately prior to tissue sampling. Only fish from the representative ancestral state (i.e. marine; '*' in red) were used as broodstock for estimating quantitative genetic parameters. Among-population comparisons also included F_2 individuals produced from a full-sib breeding design, indicated by the thin, parallel green and blue lines.
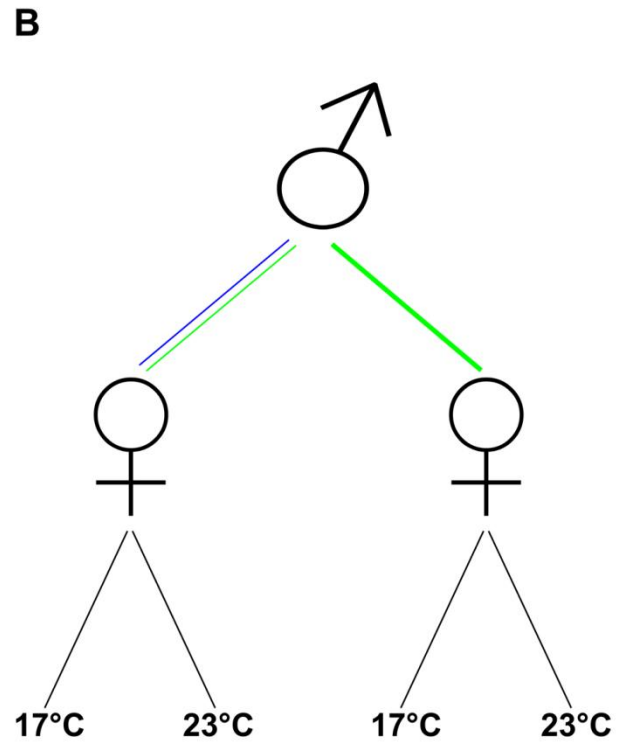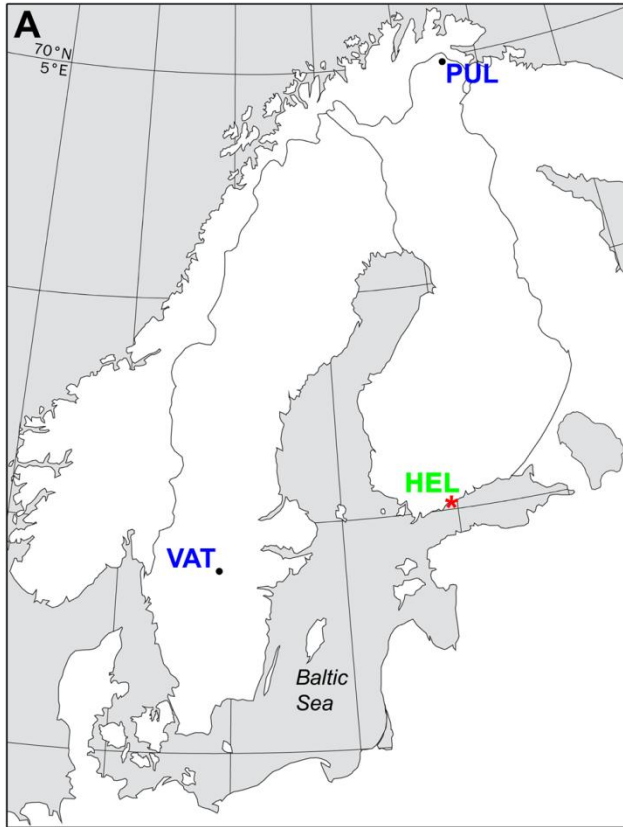
FIG. S2. Realized power and bias of the experimental pedigree. (**A**) Probability of detecting significant heritability of gene transcription with the pedigree used in the quantitative genetic analyses, based on phenotypic simulation. The full range of 'true' heritability values used in simulations is shown. Estimated power (solid line) is bounded by upper and lower 95% confidence limits (broken lines). (**B**) A magnified view of that portion of the power curve (green box in panel A) corresponding to the range of heritability values with potentially problematic detection – this includes those values with low power of detection, and the estimated false detection rate (i.e. 265 "significant" estimates of 1,000 simulated phenotypes with $h^2 = 0$). (**C**) Comparison of heritability estimates and expected/true values. Point estimates, averaged over 100 simulations per 'true' heritability value, are plotted as a solid black line; broken lines show the upper and lower 95% posterior density intervals for the estimates. The red line denotes a 1:1 relationship. (**D**) False estimates of d^2 from data simulated under a range of heritabilities & $V_D = 0$ (100 simulations per h^2 interval). Barplot height represents the mean of point estimates from simulations in which V_D was erroneously deemed signficiant (h^2 in green & d^2 in red); error bars denote the mean of upper and lower 95%PDIs.
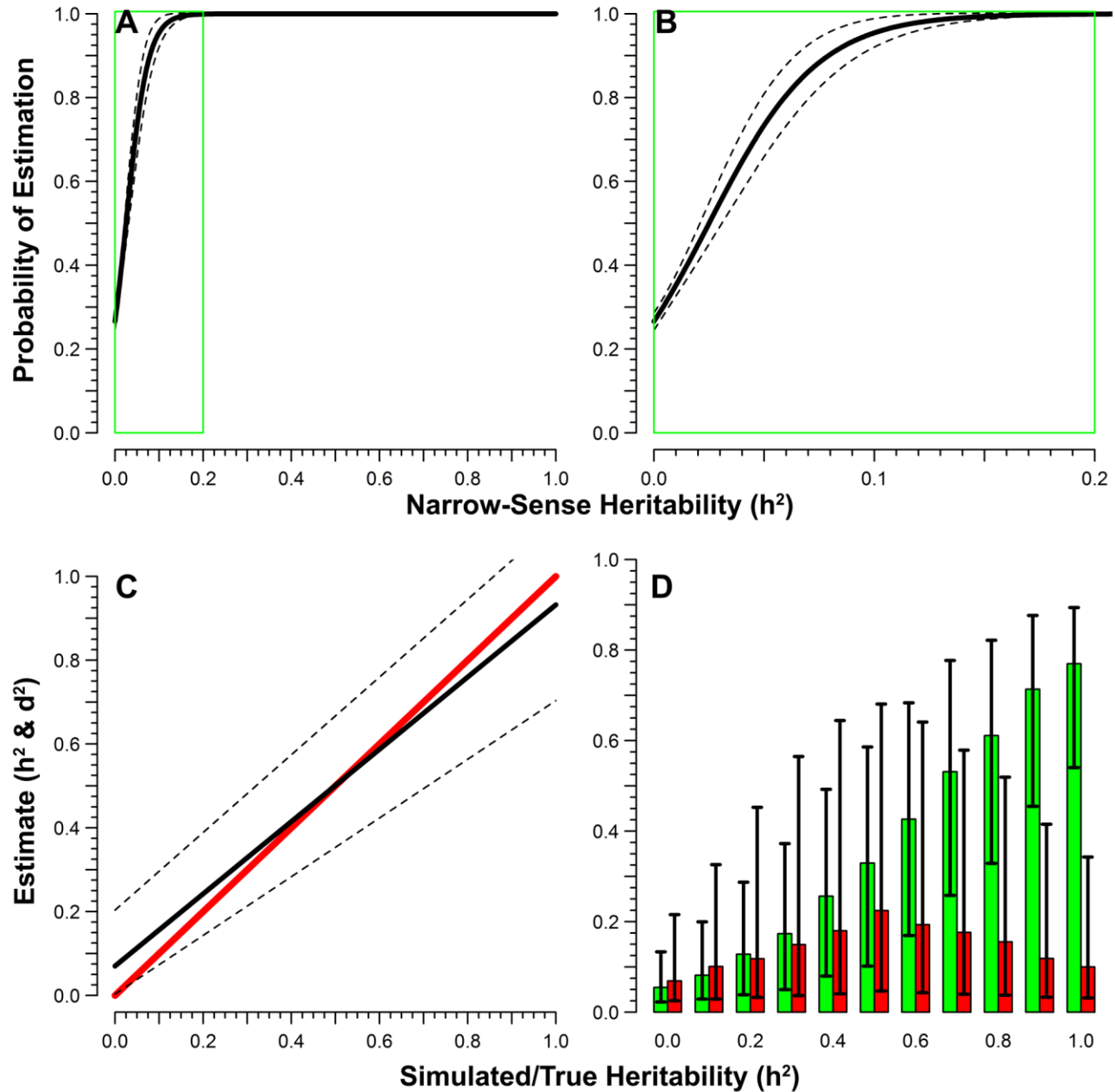
FIG. S3. Among-population divergence in transcript abundance plotted as a function of position for chromosomes XIX (**A**) and XXI (**B**). Loess smoothing (5kb intervals) of lower 95% posterior density interval values indicates putative genomic regions rich in adaptive expression divergence. Green triangles denote point estimates of Q_{ST} for transcripts which significantly exceed a baseline of neutral divergence, as defined by 17 microsatellite markers (range defined by red horizontal lines; see Methods for details). Probes whose levels of expression divergence fell within the range of neutral differentiation are shown in open, red circles.
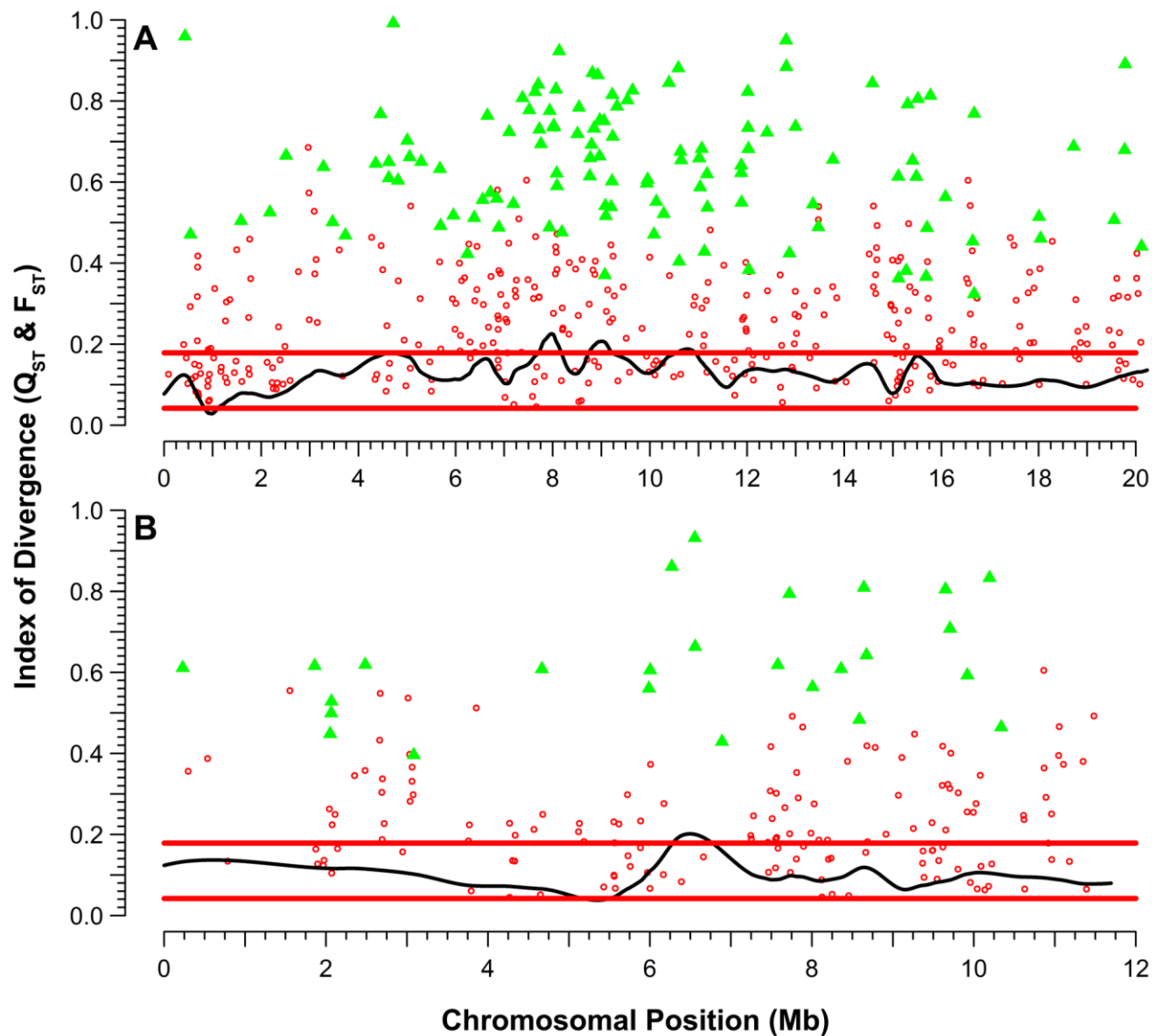
FIG. S4. Genetic proportions of total phenotypic variance. (**A**) Frequency distribution of point-estimates of the cumulative genetic proportion of total phenotypic variance (G^2) for all 10,527 transcripts. (**B**) Scatter plot of the relationship between G^2 and the absolute magnitude of differences between phenotypic variance estimates in each temperature treatment (ΔV_P), relative to total phenotypic variance (V_P). The 6,987 transcripts with a significant temperature response are plotted. (**C**) Loess smoothing of G^2, plotted as a function of genomic location (5kb intervals). Trends in point estimates are presented in orange, bounded by 95% PDIs (thinner, black lines). (**D**) Comparison of trends in G^2 (orange) and ΔV_P as a proportion of total V_P (blue), plotted as a function of genomic location. The cumulative means of these estimates are plotted in mauve.
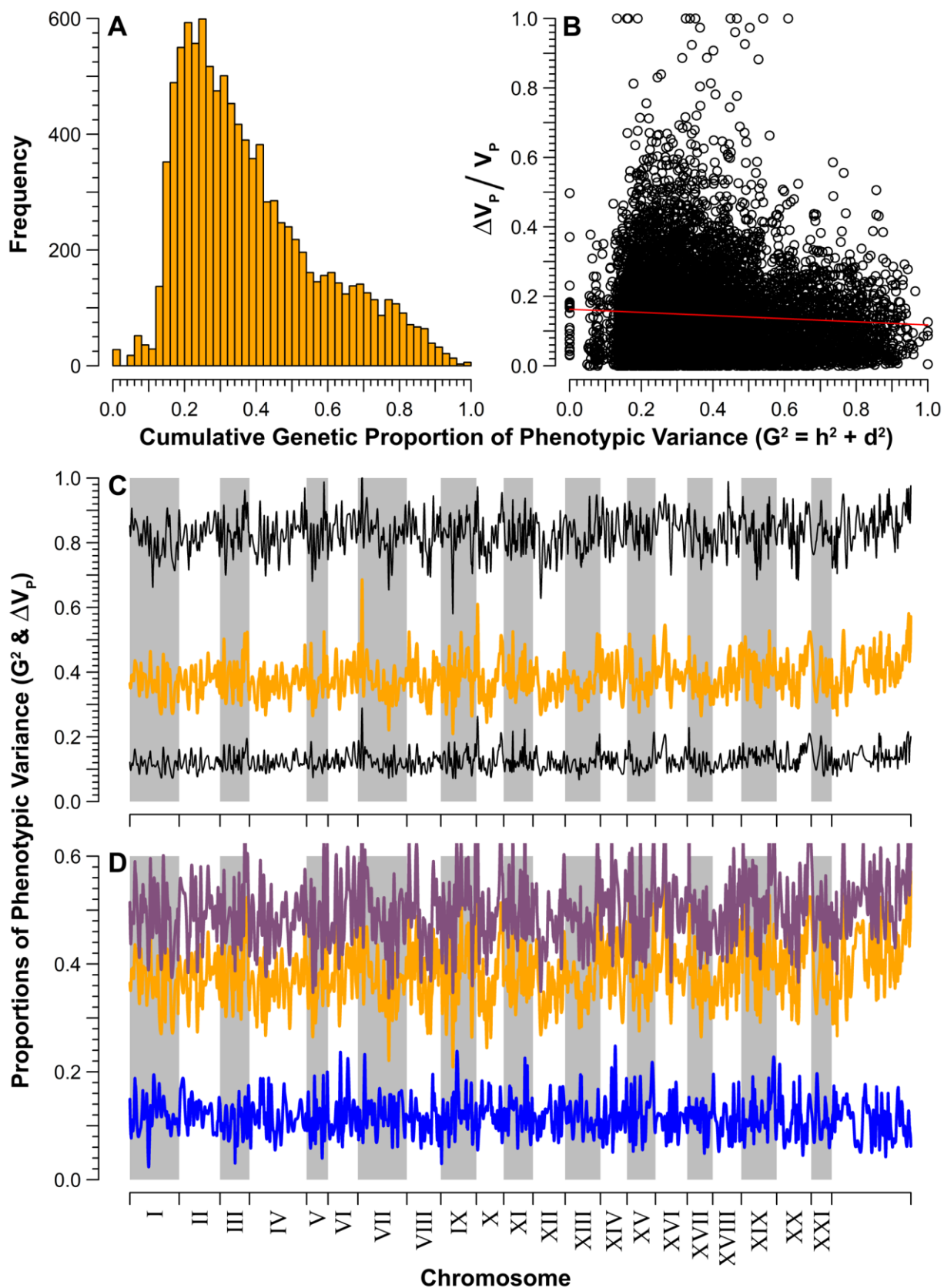
FIG. S5. Contrasting inferences based on rates of divergence (Δ) and Q_{ST}. (**A**) Overlap between sets of transcripts. In both analyses, significance of directional selection (Dir. Sel.) and neutrality was inferred on the basis of exclusion or overlap of 95%CIs with a neutral region. (**B**) Overlap between Q_{ST}-based inference and a common/typical analysis in which inference is based on only a single point estimate of Δ. Under less stringent significance criteria stabilizing selection (Stab. Sel.) is detected. Note that the region describing neutrality based on Δ (not shown) overlaps with that of the neutral Q_{ST} region, but also includes the 1,143 transcripts inferred to have experienced directional selection in the Q_{ST} analysis (Dir. Sel. Q_{ST}). (**C**) Magnitude of difference between upper and lower confidence limits ($CI_{0.975}/CI_{0.025}$) for Δ estimates. The shaded region (rose coloured) denotes neutral parameter space, and the horizontal line (red) denotes the magnitude of difference for this region. Transcripts inferred to be under directional selection, based on Δ CI exclusion, are plotted as triangles (cyan); note that point-estimates for an additional 359 transcripts appear to exceed neutral expectation, but were inferred non-significant on the basis of CI overlap. Transcripts appearing to be under stabilizing selection, based only on point-estimate exclusion, are plotted in pale yellow – none are significant using a strict criterion of CI exclusion. (**D**) Magnitude of difference of PDIs for Q_{ST} estimates. Transcripts inferred to be under directional selection are plotted in triangles (green). Transcripts under stabilizing selection, inferred by Δ, are plotted in pale yellow. Shading and horizontal line as in panel C.