SUPPLEMENTARY FIGURES

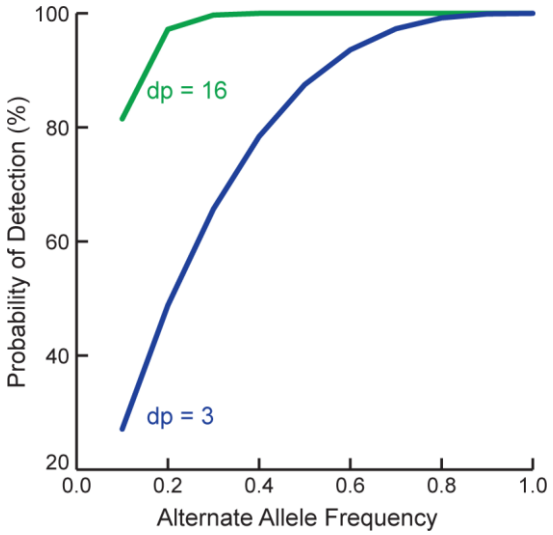

Figure S1. **Expected SNP Detection**. A scatter plot demonstrates the probability of sequencing an alternate allele (y-axis) given the alternate allele frequency (x-axis) within our sample at the average (16 reads – green) and minimum (3 reads – blue line) read depths, respectively. % Detection = $[(1\text{-AF})^{dp}]*100$ Abbr: AF=alternate allele frequency; dp – read depth.

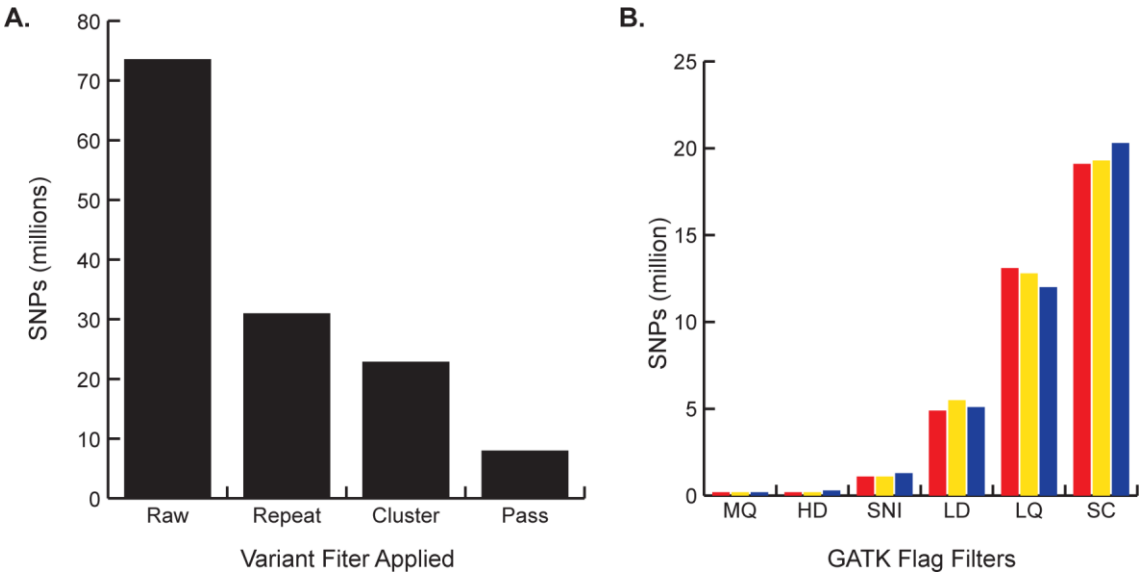

Figure S2. Filtering of SNP Variants in Laboratory Zebrafish. **A.** A bar graph showing the stepwise filters (left to right) applied to the overall number of SNPs (FLI, TL, & WIK combined). Raw refers to the total number of SNPs identified by GATK. Repeat represents the number of SNPs after excluding repetitive positions in Zv9 annotated by Repeatmasker. Cluster denotes the number of SNPs that have > 3 reads in all of the surveyed strains (FLI, TL, and WIK). Pass indicates the number of SNPs that were not flagged by the GATK. **B.** A bar graph depicting the performance of each GATK flag filter after removal of repeatmasked variants. The following abbreviations have been used for each filter: MQ = map quality; HD = high depth; SNI = SNP Near Indel; LD = Low Depth; LQ = low quality; SC = SNP clustering. Data from each strain is shown as FLI (red), WIK (yellow) and TL (blue).

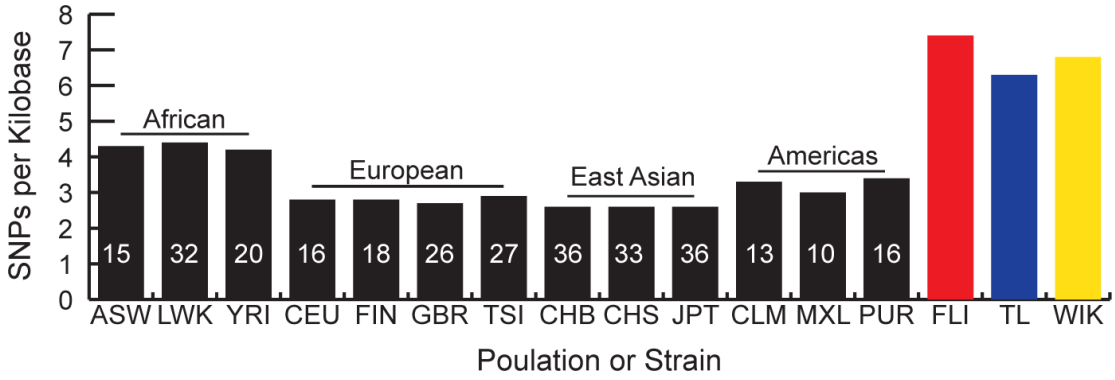

Figure S3. **Variant Genome Positions of Man and Laboratory Zebrafish.** A bar graph demonstrating abundance of SNPs (SNPs per kb of unique sequence) in the sampled zebrafish strains [FLI (red), TL (blue), WIK (yellow) to equivalent population samples constructed by random selection of 18 genomes within human populations analyzed by the 1000 genome project (31). The numbers within the bars indicate the number of calculations performed for each human population. Abbr: ASW = people with African ancestry in the Southwest United State; CEU = Utah residents with ancestry from Northern and Western Europe; CHB = Han Chinese in Beijing, China; CHS = Han Chinese in South China; CLM = Colombians in Medellin, Colombia; FIN = Finnish in Finland; FLI = *Tg(fli1a-eGFP)*^{y1}; LWK = Luhya in Webuye, Kenya; JPT = Japanese in Tokyo, Japan; MXL = people with Mexican ancestry in Los Angeles, California; PUR = Puerto Ricans in Puerto Rico; TL = *Tüpfel Long Fin*; TSI = Toscani in Italia; YRI = Yoruba in Ibadan, Nigeria.

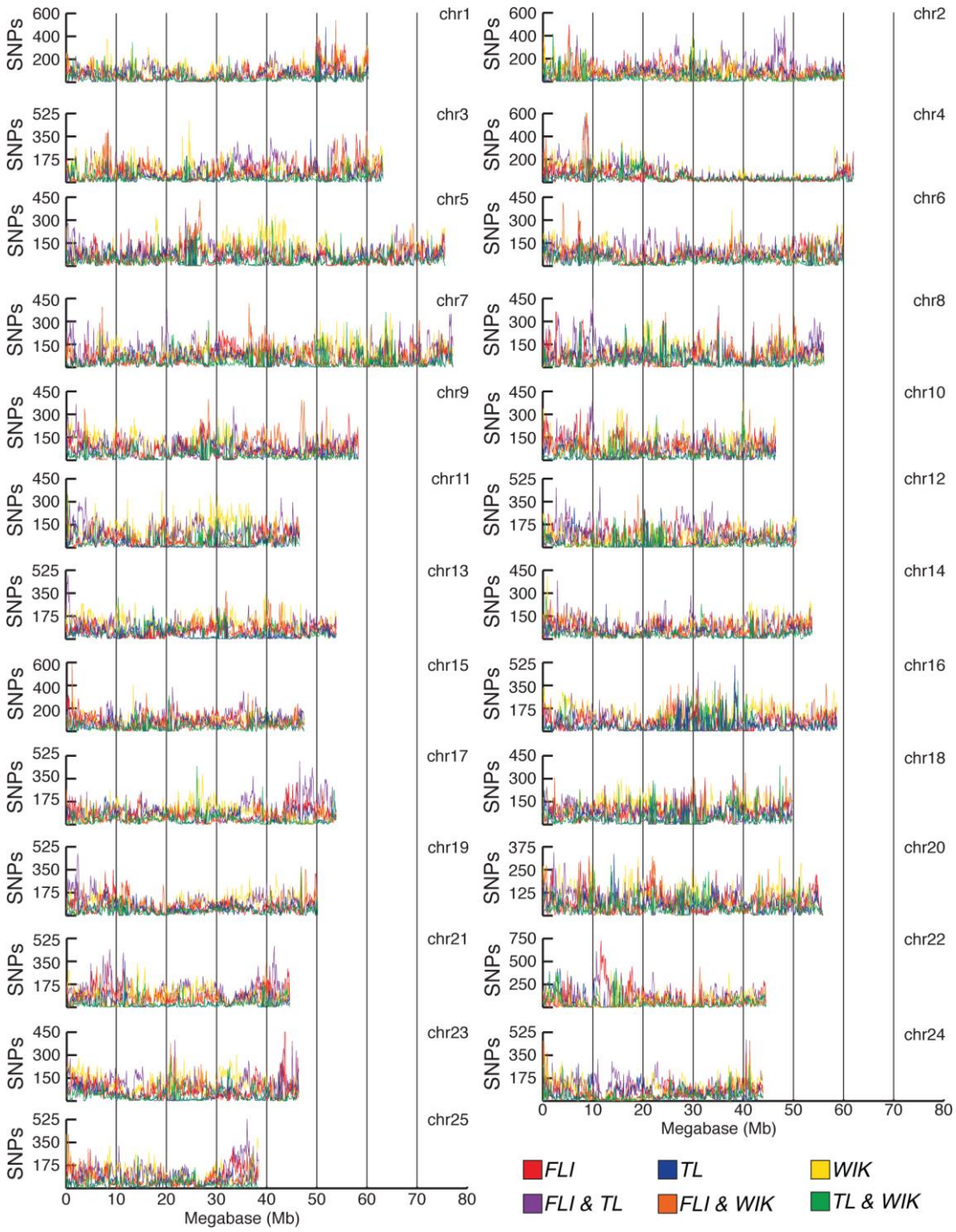

Figure S4. Spatial Distribution of Strain Variation in Laboratory Zebrafish. Each zebrafish chromosome was divided into 100kb non-overlapping windows and the number of strain variants were counted. A scatterplot was generated for each chromosome with the chromosomal position (in megabases on the x-axis) versus the number of strain variants (y-axis). Variants were color-coded corresponding to their observed strain distribution (FLI = red, TL = blue, WIK = yellow, FLI&TL = purple, FLI&WIK = orange, TL&WIK = green, and ALL = black).

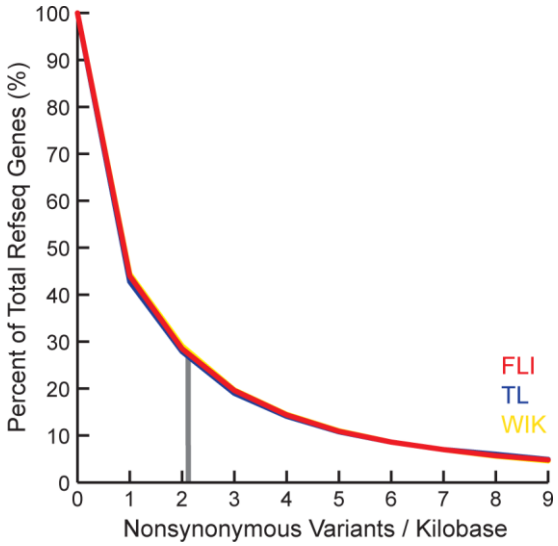

Figure S5. Non-synonymous Mutations Per Refseq Gene. A scatter plot of the minimum number of non-synonymous mutations per kb (x-axis) compared to the percentage of the total number of Refseq genes (14,484 genes). The number of non-synonymous mutations is expressed as the number of mutations per kb to prevent bias based on gene size. The mean number of nonsynonymous mutations per kb (2.1) is represented by the gray vertical line.