Supplementary Information

Identifying the Candidate Gene for HSCV by RNA-seq analysis

We performed RNA-seq analysis to identify genes implicated in HSCV. In addition to the typical RNA-seq analysis, we also performed SNP analysis in parallel to narrow down the list of candidate genes. Four representative strains have been used in this study: FGSC 8831^{AF} (High RI), FGSC 8578^{NA} (Low RI) parents and their progeny, an isogenic African population N309 with a High RI phenotype 309-89^{af} and a North American population with a Low RI phenotype 309-50^{na} (*SI Appendix*, Fig. S6).

The processed reads were mapped to the reference genome from JGI (https://jgi.doe.gov/) using TopHat2 (v2.1.1) (Kim et al., 2013). HTSeq framework (v0.9.1) (Anders et al., 2015), was used to count the aligned reads in genes. Mode "union" and a mapping quality cutoff of 20 were used for our analysis. The count-table was normalized so that all samples had the same level of total mapped reads. Samtools mpileup package (Li, 2011) was used to call SNPs and Indels. SNPs identified in samples belonging to the same strains were merged by summing up the counts of A, T, G, or C. The 4 strains: FGSC 8831^{AF}, FGSC 8578^{NA}, 309-89^{af}, and 309-50^{na} were noted as 1, 2, 3, 4 sequentially. The SNP distance between pairs of strains noted as D_{12}, D_{13}, D_{14}, D_{23}, D_{24}, and D_{34} was defined by the following equation 1:

$$D_{ik} = \sum_{1<i<k<4} \sum_{j\in(A,T,G,C)} (R_j^i - R_j^k)^2$$

R is the SNP ratio. R_i represents the R of strain i, and R_k represents the R of strain k. i and k belongs to strain 1, 2, 3, and 4. For example, D_{12} is the distance by equation 1 between strain 1 (FGSC 8831^{AF}) and Strain 2 (FGSC 8578^{NA}).

The SNPs were selected based on the following criteria: the read depth coverage of the SNP in all 4 strains must be >10, and the summation of all 6 distances must be >1. The distance ranged between 0 and 2. A pairwise distance of 0 means that the two strains have the same genotype at this SNP; a distance of 2 means that the two strains have different genotypes at this SNP. Based on the 6 pairs of distances, 64 expected genetic models were constructed (*SI Appendix, Fig. S7*). $\dot{D}$ is defined as the vector for (D_{12}, D_{13}, D_{14}, D_{23}, D_{24}, and D_{34}). We assumed that the SNPs between

2 strains are ideally the same or different, and then the expected distance is either 0 or 2. The 0/2 combinations of 6 pairwise distances comprise the 64 models. $\dot{D}_{mi}$ is denotes the vector of 64 models, and mi is the serial number of the models. For example, m30 means model 30 in *SI Appendix*, Fig. S7. $\dot{D}_{SNP}$ denotes the empirical values of the distance vector. The distance between each selected SNP and the 64 expected models is defined as:

$$D_{SNP-mi} = (\dot{D}_{SNP} - \dot{D}_{mi})^2$$

EdgeR (Robinson et al., 2010) and DEseq (Anders and Huber, 2010) were used to calculate deferentially expressed genes (DEG). The DEG comparison was performed for: FGSC 8831^{AF} and FGSC 8578^{NA}; $309\text{-}89^{af}$ and $309\text{-}50^{na}$ at each time point, 15M (15 min after dark to light transition), 1H (1 hour after dark to light transition), 2H (2 hours after dark to light transition) and 2DK (2 hours darkness).

Based on the specific criteria (SNP read depth >10 and summation of all 6 distance >1), 120,314 SNPs were retained from the mapped data of 48 samples from 4 strains (three experimental replicates with time points in light (15 min, 60 min, 2 hours) and dark (2 hours) of 12-hour:12-hour LD cycle). These SNPs were further assigned to 64 models based on the criteria of $D_{SNP\text{-}mi} < 0.01$. We derived these 64 models to explain how genetic information is passing between parents and progeny strains. Except for an association of one SNP, which was assigned to model no. 15, all other SNPs were assigned to model numbers 20, 30, 38 and 44 (*SI Appendix*, Fig. S7), and SNPs associated with these 4 models segregate with the *N. discreta* genome (*SI Appendix*, Fig. S8). Among these four models, phenotype specific genetic inheritance from parents to progeny has been clearly observed with model number 30. We additionally plotted a heat-map with model no. 30 DEGs, and identified a a functionally annotated candidate gene using DEGs in all time points of the light phase of the 12-hour cycle: 12-hour LD cycle. This candidate gene, NEUDI_158280, also displays association with conidia formation among High and Low RI phenotypes. No significant variation of gene expression is reported in Low RI phenotypes between light and dark phases of the 12-hour LD cycle.

Based on the allelic and expression variation of NEUDI_158280, we further understand the behavior of NEUDI_158280 in high and low RI phenotypes in 24-hour light-dark cycles. Our

qPCR results suggest that NEUDI_158280 is rhythmic in diurnal conditions in High RI phenotypes (Circwave, $p < 0.001$) compared to Low RI phenotypes (Fig. 4*A*), and the relative expression of NEUDI_158280 is significantly lower in Low RI compared to High RI phenotypes. To strengthen our hypothesis about association of NEUDI_158280 with HSCV, we did a segregation analysis of NEUDI_158280 against multiple strains from the N309 population, which includes High and Low RI phenotypes. We observed segregation in mRNA levels of NEUDI_158280 between Low and High RI (Fig. 4*B*), which has not been observed for the FRQ protein (Fig. 2*D*). mRNA levels of NEUDI_158280 are higher in High RI compared to Low RI phenotypes. These experiments suggest that NEUDI_158280 plays a potential role in habitat specific clock variation of *N. discreta Ps4b* natural populations.

Table S1: Primers for qPCR validation

Ascension Number	Description	Primer Sequence (5'-3')	Amplicon Length	T_m	PCR Efficiency	Final Conc.
NEUDI_98553	Beta-Tubulin	F: CCACTTCTTCATGGTCGGCT R: CCTAGGGTCGAACATCTGCT	103	61.7 57.6	2.08	500nM
NEUDI_168658	Vacuolar ATPase subunit 2 (*homolog of N. crassa*)	F: GTCGTCCAGGTCTTTGAGG R: TGCCGGAACCATCGAAGATACG	127	59.1 64	2.1	500nM
NEUDI_158280	Ubiquinol cytochrome c reductase, subunit RIP1	F: GCGACAAAGACCTCAAGACC R: GTTCGTAGGCGGTCGAAATA	72	60.2 59.1	2.18	500nM

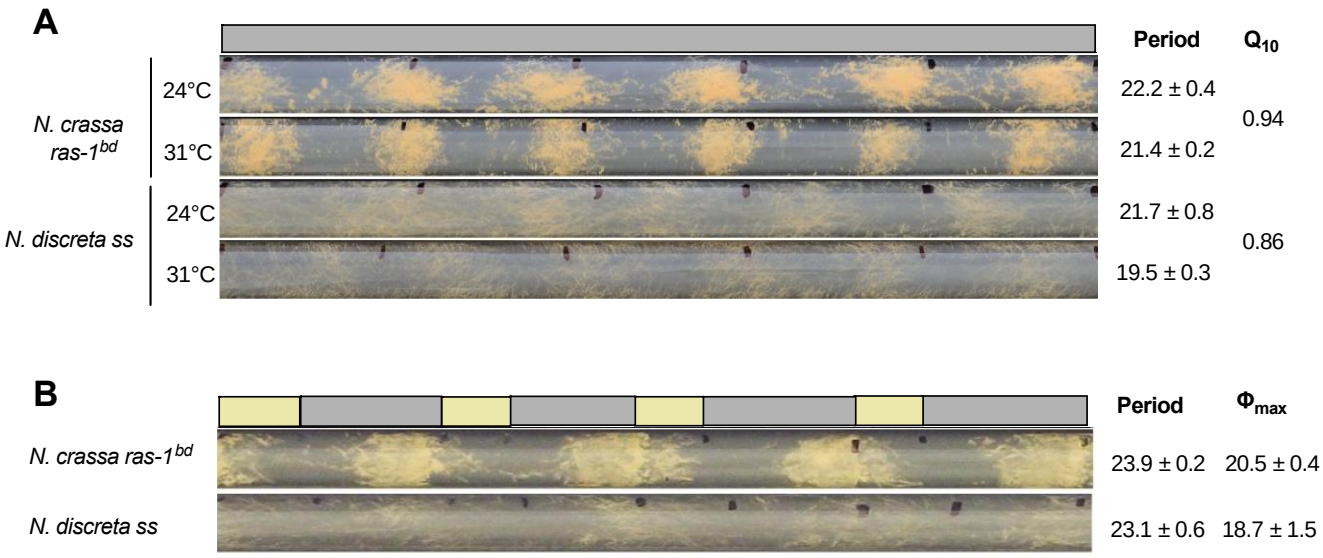

Fig. S1. Asexual development is circadian regulated in *N. discreta sensu stricto*.
(A) Race tube analysis of *N. crassa ras-1*^{bd} and *N. discreta sensu stricto* under constant darkness at 24ºC and 31ºC. Free-running periods (FRP) for each strain (n=6) are shown with standard deviations. Temperature compensation of each strain was assessed by the temperature coefficient (Q_{10}) using the following equation: $Q_{10} = (FRP^{31ºC}/ FRP^{24ºC})^{10/(31ºC-24ºC)}$. (B) Twelve-hour light and dark cycles (LD 12:12) were used to test for photic entrainment of the overt rhythms at 26.2ºC ± 0.7ºC. The mean light intensity was 4,261 ± 1,472 lux with an average ambient temperature of 26.7ºC ± 0.5, whereas the light intensity was null under dark phase and the mean temperature was 25.6ºC ± 0.0. Entrained phase values indicate when the spore reaches its max during the 24h cycle (ϕ_{max}) in zeitgeber time (ZT). Under ZT time, hour 0 and 24 are the same and this time point is by definition the moment when lights on.

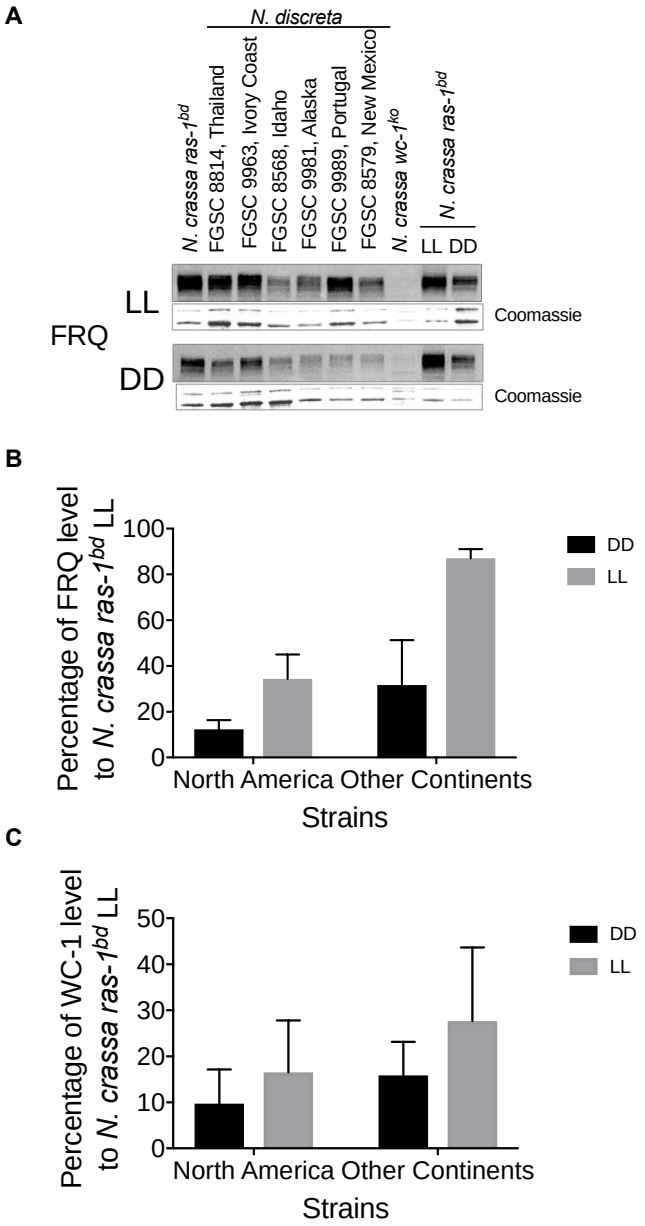

Fig. S2. Segregation of FRQ and WC-1 protein levels in North American strains of *Neurospora*. (A) western blot analysis of FRQ and WC-1 protein in representative *N. crassa* and *N. discreta* strains in constant light (LL) and constant dark (DD). (B) Relative levels of FRQ protein in *N. discreta* from North America and Other Continents, relative to *N. crassa ras-1*^{bd} (from LL) in LL or DD. (C) Relative levels of WC-1 protein in *N. discreta* from North America and Other Continents, relative to *N. crassa ras-1*^{bd} (from LL) in LL or DD.

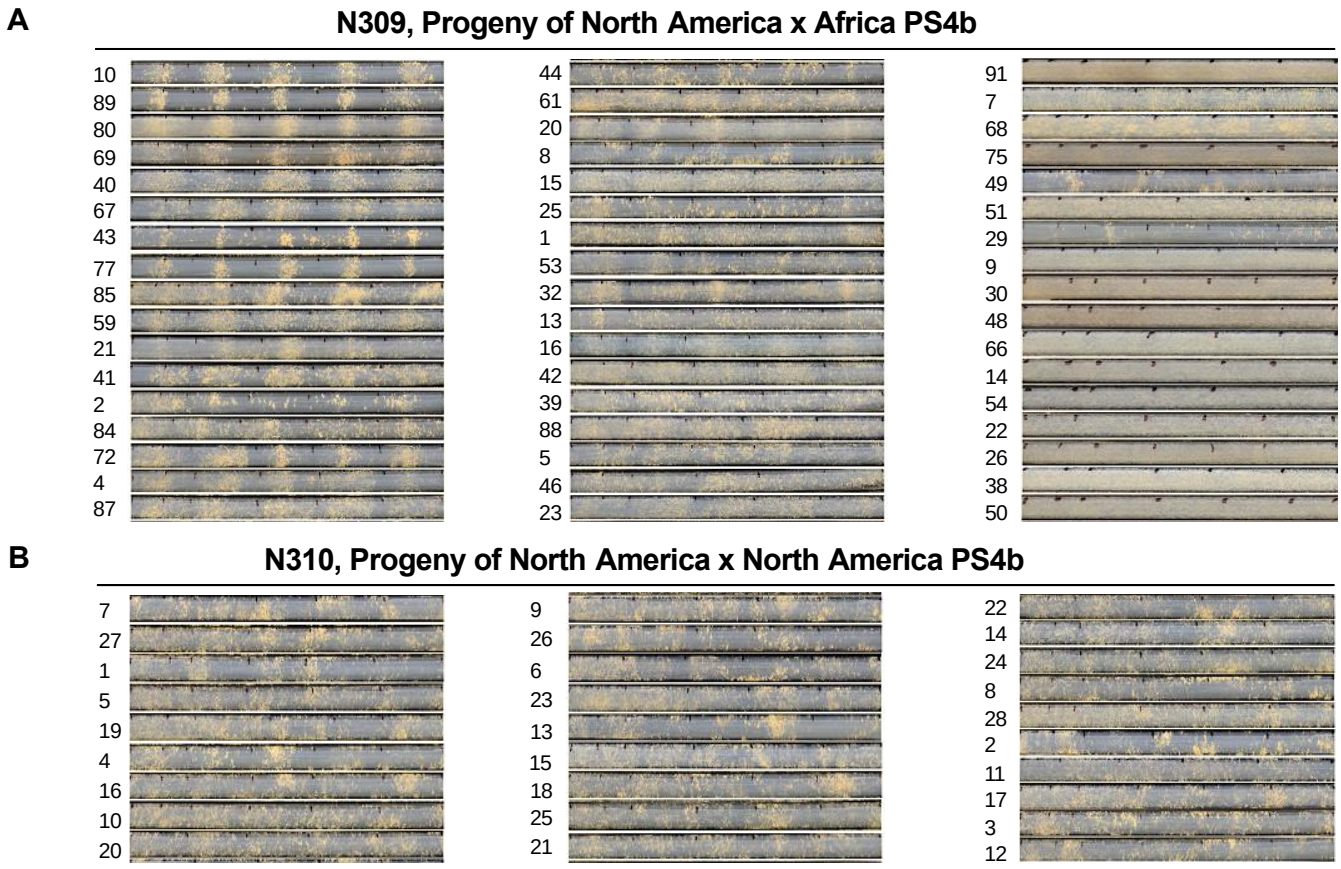

Fig. S3. Quantitative rhythmicity index (RI) phenotype in F1 progenies.
Two populations were created to test if the banding phenotype was caused by quantitative or Mendelian genetics. (A) An inter-continental population was created using the strains FGSC8578 New Mexico, USA and FGSC8831, Ivory Coast, Africa. Strains are depicted from highest RI (top left) to lowest RI (bottom right). Numbers indicate different strains within the population. (B) An intra-continental population was created using two strains from New Mexico, USA: FGSC8579 and FGSC8578. This was used as a control to determine if the banding phenotype is directly from Africa's HSCV and not an artefact of crossing.

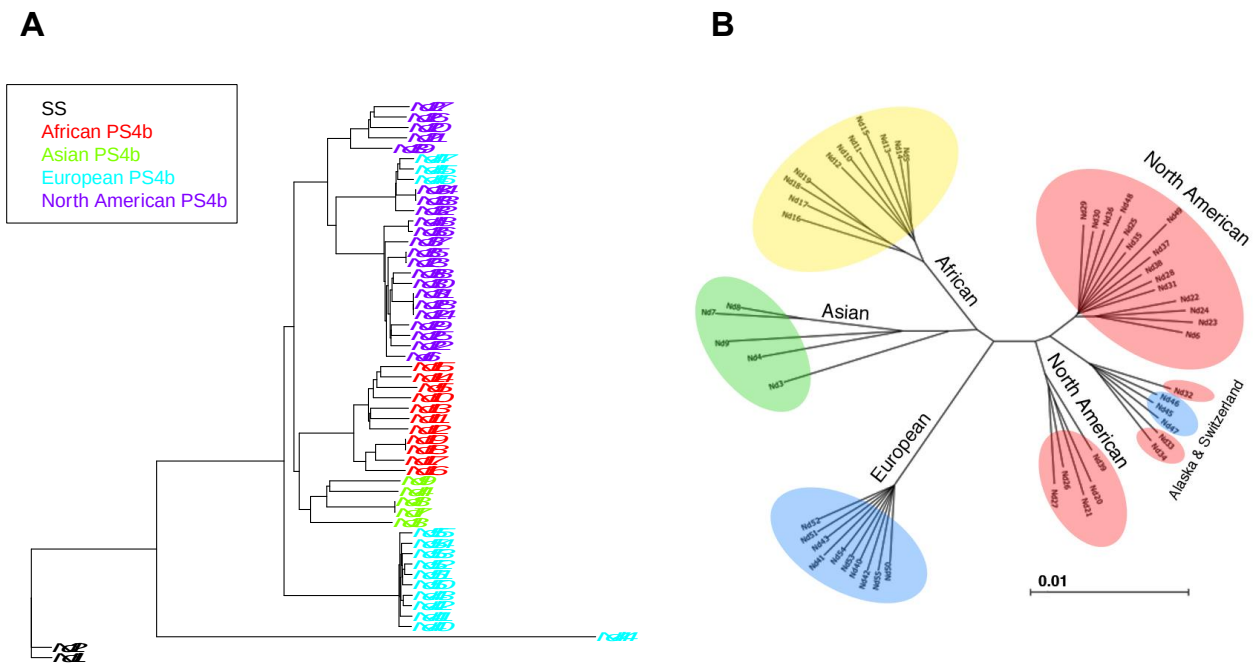

Fig. S4. Phylogenetic analysis of PS4b strains.
(A) Rooted tree. (B) Un-rooted tree. The oldest split is between Nd44 and SS strains (Nd1 and Nd2). The 2nd oldest split is the European strains split (excluding Nd45, Nd46, Nd47). The 3rd oldest split is between the North American strains (plus Nd45, Nd46, Nd47) and the Asian/African strains. The 4th oldest split is between African and Asian strains. North American strains (plus Nd45, Nd46, Nd47) have been separated from the other continental strains for about as long as the African and Asian strains have been separated from European strains (excluding Nd45, Nd46, Nd47). Based on the analysis, we hypothesized that Nd45, Nd46, Nd47 may have originated in North America and then at some point very recently traveled back to Europe.

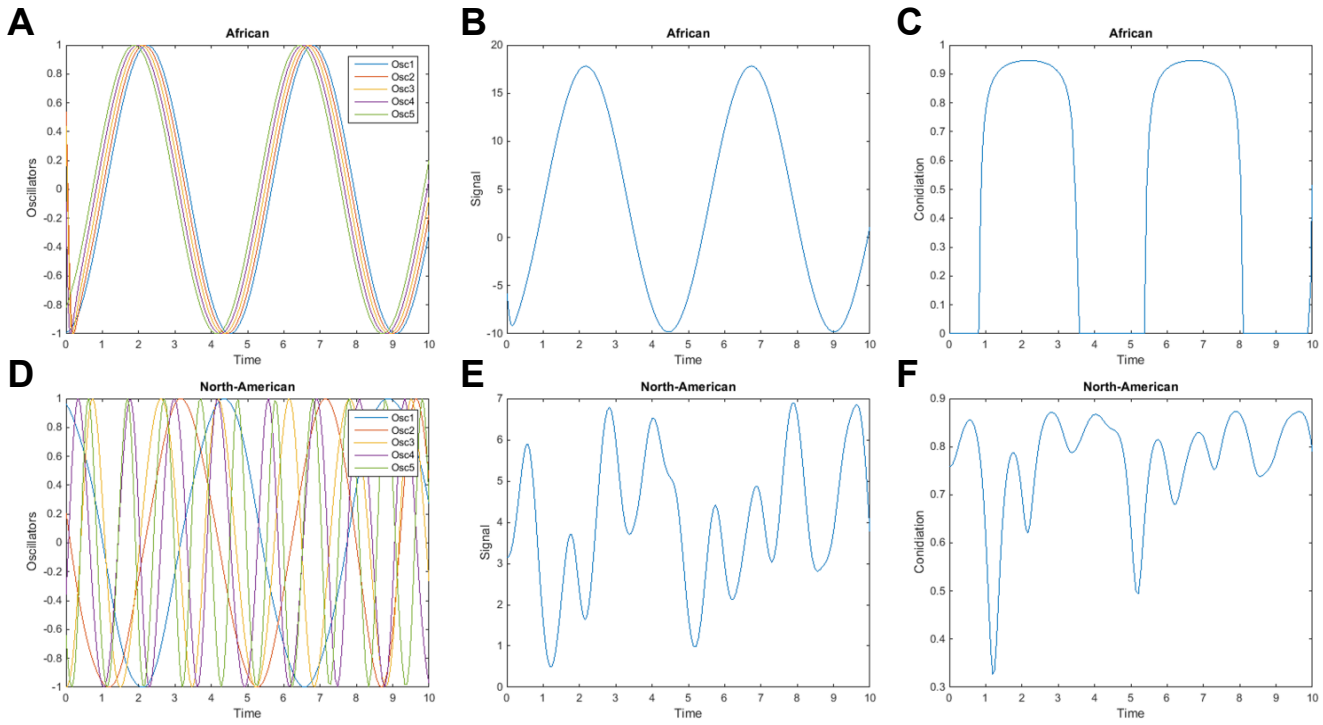

Fig. S5. Mathematical model of fitness. (A) Time evolution (T=10) for n=5 coupled oscillators with strong coupling (K=10). Rapidly all oscillators are synchronous with small phase shift. (B) Total signal for (a). (C) Conidiation for (a). (D) Time evolution with weakly coupling (K=1). Oscillators remain uncoupled. (E) Total signal for (d). (F) Conidiation for (d).

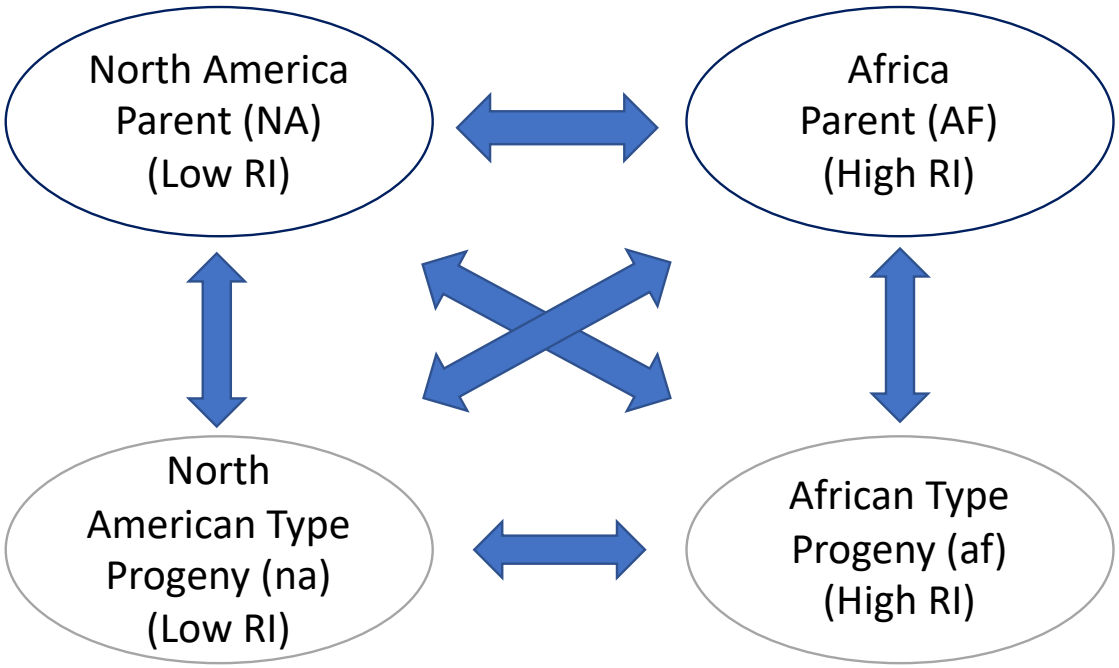

Fig. S6: Model for identifying nucleotide similarities/ dissimilarities between African and North American parents and progeny. The left panel of the model represents North American parent and progeny. The right panel represent the African parent and progeny. The arrow marks indicate the comparisons used for the identification of similarity/dissimilarity of nucleotides between parents and progeny of North American and African strains.

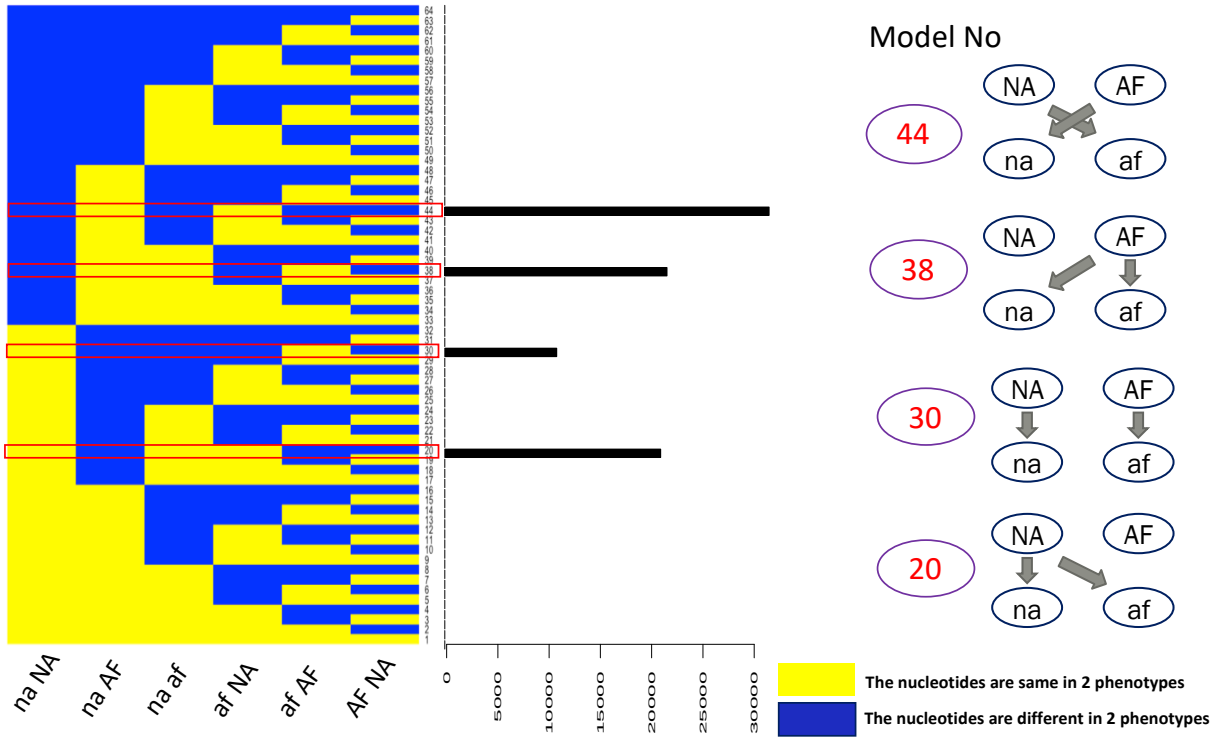

Fig. S7: Nucleotide similarities and dissimilarities between Parents and Progeny: Yellow and Blue represent similarities and dissimilarities of nucleotides in combinations of parents and progeny, respectively. African and North American parents are represented as AF and NA, respectively. Progeny of African and North American types are represented as af and na. The middle panel represents the number of SNPs identified in the RNA sequencing data, which belongs to four models (20, 30, 38 and 44). The right panel represents the respective models.

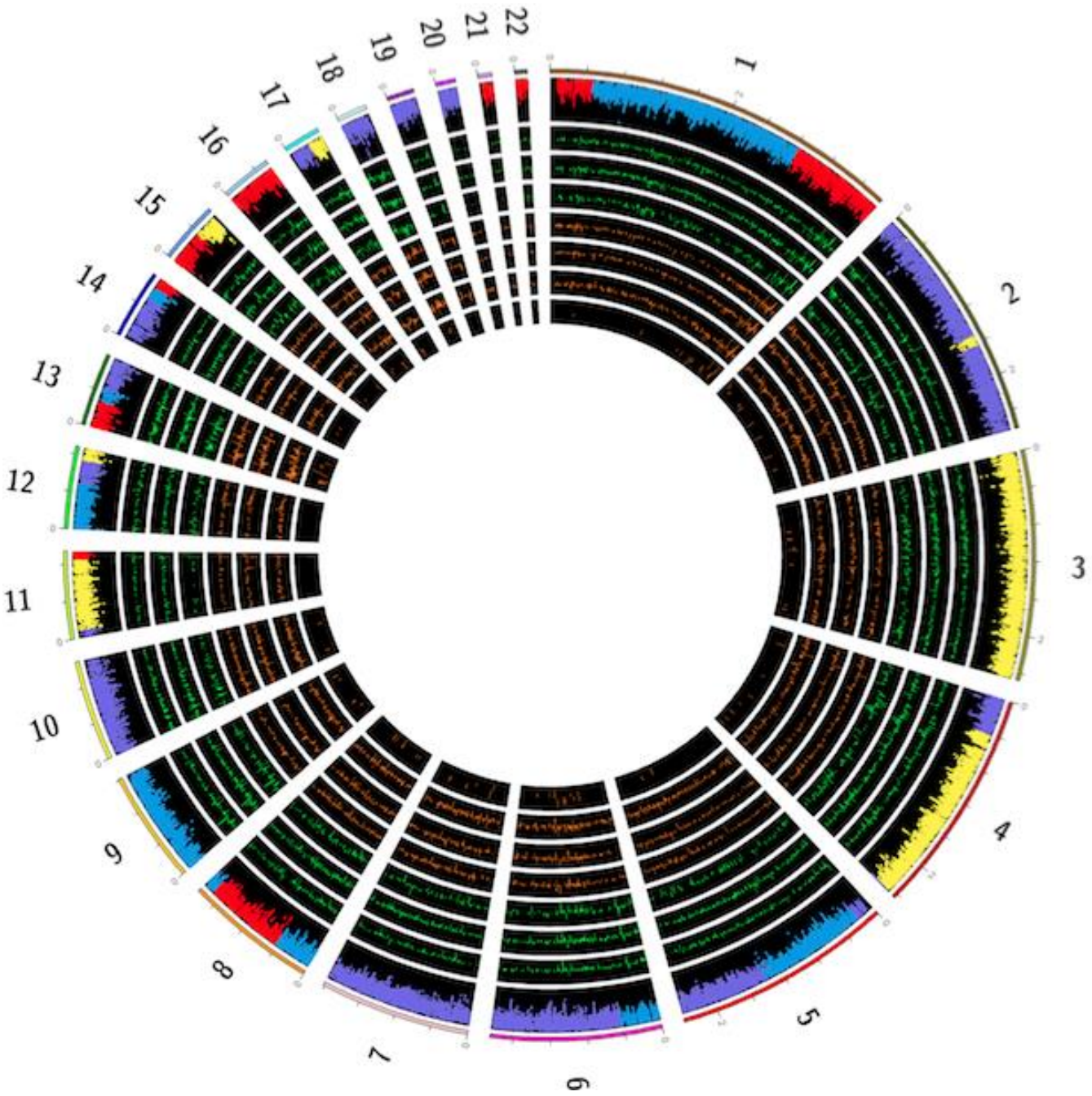

Fig. S8. The scatterplot in 1st layer from outside is $\log_{10}$ depth of SNPs of 4 models. Model 20 is yellow; Model 30 is red; Mode 38 is blue; Model 44 is purple. The logs FC of differentially expressed genes are shown from layer 2 to 8 in green or orange histograms. The green histograms are for log FC of the comparison between FGSC $8831^{AF}$and FGSC 8578^{NA} at 2H, 1H and 15M from layer 2 to 4. The orange histograms are for log FC of the comparison between $309\text{-}89^{af}$, and $309\text{-}50^{na}$ at 2H, 1H, 15M, and 2DK from layer 5 to layer 8.

Anders S, and Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11:R106.

Anders S, Pyl PT, and Huber W (2015) HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 31:166-169.

Edgar R, Domrachev M, and Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 30:207-210.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, and Salzberg SL (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14:R36.

Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27:2987-2993.

Robinson MD, McCarthy DJ, and Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139-140.