

***Cis*-regulatory evolution in prokaryotes revealed by interspecific archaeal hybrids**

Carlo G. Artieri^{1,4*}, Adit Naor^{2*}, Israela Turgeman-Grott³, Yiqi Zhou¹, Ryan York¹, Uri Gophna³
and Hunter B. Fraser^{1†}

¹Department of Biology, Stanford University, Stanford, CA 94305, USA.

²Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA 94305, USA

³Department of Molecular Microbiology and Biotechnology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv 6997801, Israel

⁴Present address: Counsyl Inc., South San Francisco, CA, 94080, USA.

*These authors contributed equally

†Corresponding author

CONTACT INFORMATION:

Hunter B. Fraser
Herrin Labs Rm 305
371 Serra Mall
Stanford, CA 94305
United States

TELEPHONE NUMBER: 650-723-1849

FAX NUMBER: 650-724-4980

EMAIL: hbfraser@stanford.edu

RUNNING HEAD: cis-regulatory adaptation in archaeal hybrids

NUMBER OF FIGURES: 3

NUMBER OF TABLES: 0

NUMBER OF SUPPLEMENTS: 2 files

KEYWORDS: archaea, cis-regulation, gene expression, allele-specific expression

Abstract

The study of allele-specific expression (ASE) in interspecific hybrids has played a central role in our understanding of a wide range of phenomena, including genomic imprinting, X-chromosome inactivation, and cis-regulatory evolution. However across the hundreds of studies of hybrid ASE, all have been restricted to sexually reproducing eukaryotes, leaving a major gap in our understanding of the genomic patterns of cis-regulatory evolution in prokaryotes. Here we introduce a method to generate stable hybrids between two species of halophilic archaea, and measure genome-wide ASE in these hybrids with RNA-seq. We found that over half of all genes have significant ASE, and that genes encoding kinases show evidence of lineage-specific selection on their cis-regulation. This pattern of polygenic selection suggested species-specific adaptation to low phosphate conditions, which we confirmed with growth experiments. Altogether, our work extends the study of ASE to archaea, and suggests that cis-regulation can evolve under polygenic lineage-specific selection in prokaryotes.

Introduction

For the past 50 years, interspecific hybrids have been an invaluable resource for studying the regulation of gene expression. Beginning with studies in species such as frogs and trout, the concept of allele-specific expression (ASE) was first used to investigate differences in enzyme activity levels for the two alleles in a hybrid (Wright and Moyer 1966, Hitzeroth et al. 1968). Since then, measurement of ASE in hybrids has played a critical role in the study of genomic

imprinting, X-chromosome inactivation, and cis-regulatory evolution (Avisé and Duvall 1977, Dickinson et al. 1984, Bartolomei et al. 1991, Wittkopp et al. 2004, Wang et al 2010, Pastinen 2010, Wittkopp and Kalay 2011, Wang and Clark 2014, Babak et al 2015).

Particularly since the advent of high-throughput RNA-sequencing (RNA-seq), ASE in hybrids has been a major focus for studies of gene expression evolution. In a hybrid, the two alleles of each gene are present in the same cells, and thus experience all the same environmental factors/perturbations, which makes direct comparison more meaningful than when the expression profiles of different species are compared—especially when the environments of those species are not well-controlled, such as in human studies. In addition, because the two alleles in a hybrid are exposed to all the same trans-acting factors (such as transcription factors)—which can affect gene expression levels, but cannot cause allelic bias in the absence of cis-regulatory divergence—ASE reflects only cis-acting differences between alleles, which can be a useful distinction for a wide range of questions in evolutionary biology (Wittkopp and Kalay 2011).

Despite the multitude of studies employing ASE (over 650 publications when searching “allele-specific expression” or “allele-specific gene expression” in PubMed abstracts), a limitation shared by all of them is that they have been restricted to eukaryotes. The reason for this is that prokaryotes do not undergo sexual reproduction, so generating hybrids has not been possible. As a result, our knowledge of cis-regulatory evolution in prokaryotes has lagged far behind that in eukaryotes.

However, some halophilic archaea can undergo a fusion process that can generate hybrid cells (Rosenshine et al. 1989, Ortenberg et al. 1999). This process is efficient even between different species, but the heterozygous hybrid state is unstable due to gene conversion events (Lange et al. 2011), as well as large-scale recombination events that result in homozygous

recombinants (Naor et al. 2012). We overcame this obstacle by maintaining two *different* selection markers at the same genetic locus in the two parental species. In such a condition any homologous recombination event will result in swapping one selection marker for the other, and as long as one selects for *both* markers, only heterozygous cells will be able to keep dividing, assuming no ectopic recombination occurs.

We have applied this unique system to explore cis-regulatory evolution in the genus *Haloferax*. The two species we studied were *Haloferax volcanii*, isolated from the Dead Sea in Jordan (Mullakhanbhai and Larsen 1975), and *Haloferax mediterranei*, isolated from a saltern in Alicante, Spain (Rodriguez-Valera et al. 1983). While both isolation sites were characterized by high salt concentrations, they likely differed greatly in other respects, such as concentrations of mangesium and phosphate ions, raising the possibility of lineage-specific adaptations of these species to their respective environments.

Results

We have previously shown that *H. volcanii* and *H. mediterranei* are able to efficiently mate and generate interspecies recombinants (Naor et al. 2012). In order to generate a stable *H. volcanii* - *H. mediterranei* hybrid, we needed to prevent the possibility of recombination between chromosomes, thus forcing the hybrid to retain both parental chromosomes. For that we needed to create mutants that carry two different selectable markers at the same genomic location, since the two strains are syntenic (López-García et al. 1995) (Figure 1A). We used the *H. mediterranei* strain WR646 ($\Delta trpA$ $hdrB^+$), an auxotroph for tryptophan and prototroph for thymidine (Naor et al 2012), and the *H. volcanii* strain H133 ($\Delta trpA$ $\Delta hdrB$), an auxotroph for tryptophan and

thymidine (Allers et al. 2004). H133 was then modified by inserting the *trpA* selectable marker into the *hdrB* locus to generate UG241 (*trpA*⁺ Δ *hdrB*). This was done by transforming H133 with pTA160-*trpA* and selecting on media lacking thymidine, thus selecting for a double crossover event copying the *trpA* selectable marker into the *hdrB* locus. To create the stable hybrid, WR646 and UG241 were mated and colonies were selected on media lacking thymidine and tryptophan (Figure 1B and Methods).

We performed both RNA- and DNA-seq on two independently derived *H. volcanii* \times *H. mediterranei* hybrid cultures (hereafter replicates 1 and 2). Reads were mapped to a reference containing both parental genomes, and species-specific gene-level expression was calculated in reads per kilobase per million mapped reads (RPKM). Integrating ortholog and operon predictions (Price et al. 2005; DeLuca et al. 2012), resulted in 1,954 orthologous transcriptional units, hereafter referred to as ‘orthologs’, corresponding to 1,507 individual genes and 447 operons (Supplemental File 1; see Methods).

As *Haloferax* species are highly tolerant of both intra- and inter-chromosomal and plasmid copy number variation (Breuert et al. 2006), we used the DNA-seq data to identify large-scale amplifications. As expected, the ratio of plasmid coverage to chromosomal coverage varied among species and between replicates (Supplemental Fig. 1). Consequently, we restricted our analysis to orthologs found outside of amplified regions and on the main chromosomes of the two parental species, resulting in 1,526 orthologs for analysis (Supplemental Table 1). We observed similar patterns of expression levels and ASE ratios in the two biological replicates (Supplemental Fig 2).

Differential expression of the two species’ alleles within a common *trans* cellular background, known as allele-specific expression (ASE), indicates divergence of *cis*-regulation

between orthologs (Pastinen 2010, Wittkopp and Kalay 2011). In order to detect significant ASE, we employed a method that takes into account both gene length and base-compositional differences between parental alleles (see Methods) (Bullard et al. 2010; Artieri and Fraser 2014). 929 orthologs showed significant ASE at a 5% false-discovery rate (FDR), indicating the presence of substantial cis-regulatory differences between the two parental species (Figure 2A). We found no significant difference in the number of genes favoring either species' allele (453 vs. 476 favoring the *H. mediteranei* vs. *H. volcanii* allele, $\chi^2 = 0.569$, 1 degree of freedom, $p = 0.451$), suggesting that ASE was about equally likely to favor either allele.

We also tested the accuracy of our classification of genes into orthologous operons by asking whether adjacent genes within operons showed greater similarity in ASE ratios than adjacent, independently transcribed genes. Indeed, genes within operons had a significantly smaller median absolute \log_2 differences in ASE values than those outside of operons in both biological replicates (Figure 2B; Kruskal-Wallis test $p < 10^{-15}$ for each replicate). These differences may be conservative, since any errors in the operon predictions (Price et al. 2005) would lead us to underestimate their magnitudes.

Although ASE data reveal genome-wide patterns of cis-regulatory divergence, these might mostly reflect random changes due to genetic drift of neutral alleles. To identify those changes driven by lineage-specific natural selection, we and others have developed a “sign test” that detects selection acting on the regulation of entire groups of functionally related genes (reviewed in Fraser 2011). This test has been successfully applied to fungi, plants, and metazoans (Fraser et al. 2010, Bullard et al. 2010, Fraser et al. 2011, Fraser et al. 2012, Fraser 2013, House et al. 2014, Naranjo et al. 2015), but not to prokaryotes, due to the previous lack of ASE data from interspecific hybrids.

We applied the sign test to Gene Ontology gene sets from *H. volcanii* (Bins et al. 2009), to search for gene sets with ASE directionality biased towards one parental species (see Methods). Such a directionality bias represents a robust signature of lineage-specific selection (Fraser 2011). We found that genes with a known role in phosphorylation (GO:0016310) showed a significant bias in ASE directionality (ASE for 16/21 alleles favoring *H. mediterranei* in each biological replicate; permutation-based $p < 0.001$). These phosphorylation-related genes were predominantly kinases, and the “kinase activity” subset (GO:0016301) showed a similar ASE bias (ASE for 15/19 alleles favoring *H. mediterranei* in each biological replicate; permutation-based $p < 0.001$; Figure 3A, Supplemental Table 1). This suggested that genes related to phosphorylation—particularly kinases—have evolved under lineage-specific selective pressures leading to increased expression in *H. mediterranei*, or decreased expression in *H. volcanii*.

The results of the sign test led us to hypothesize that the higher expression of kinases in *H. mediterranei* may increase the fitness of this species in conditions where phosphate is limiting, since it could utilize the scarce phosphate more efficiently. To test this hypothesis, we grew both parental strains in low (0.1 mM) phosphate for 30 hours. Consistent with our prediction, *H. mediterranei* showed robust growth in this condition, in contrast to *H. volcanii*, whose growth was highly impaired (Figure 3B and Supplementary Figure 3).

Conclusion

In this work we have introduced a method to create stable interspecific hybrids of *Haloferax*, and used these hybrids to investigate the extent and phenotypic impacts of cis-regulatory evolution. Our application of the sign test revealed lineage-specific selection acting on

the cis-regulation of kinases, which led to our prediction—and confirmation—of *H. mediterranei*'s superior fitness in phosphate-limiting conditions.

Although we do not know the phosphate concentrations of the specific sites where these two species were isolated, it is well established that phosphate is the main limiting nutrient in the Mediterranean (Lazzari et al. 2016). In contrast, the Dead Sea contains higher phosphate levels, particularly in the sediments where *H. volcanii* was most abundant (Nissenbaum et al. 1990). Therefore it is not surprising that *H. mediterranei* may have adapted to the low phosphate levels with increased expression levels of kinases, compared to *H. volcanii*.

Over half (929/1526) of the orthologs we studied showed significant ASE, though this fraction would likely increase with greater sequencing depth. Based on this lower bound estimate, we conclude that cis-regulatory divergence is likely to be a major source of evolutionary novelty in *Haloferax*, and that polygenic selection can act on the regulation of groups of functionally related genes in prokaryotes, similar to patterns of polygenic adaptation that have been discovered with the sign test across a wide range of eukaryotes.

Materials and Methods

Generation of the hybrids

Strains used: WR646 ($\Delta pyrE \Delta trpA hdrB^+$), H133 ($\Delta pyrE \Delta trpA \Delta leuB \Delta hdrB$), UG241 ($\Delta pyrE \Delta leuB \Delta hdrB::trpA$). Plasmids used: pTA160 for $\Delta hdrB$ deletion in $\Delta pyrE2$ background (Allers et al. 2004), and pTA298 for making deletions in $\Delta trpA$ background (Stroud et al. 2012).

Strains were routinely grown in rich medium (Hv-YPC). When selection was needed we used Casamino Acids medium (Hv-Ca). When required, 50 μ g/ml of thymidine, uracil or tryptophan were added. Following mating all strains were grown on enhanced Casamino broth. All media were made as described (<http://www.haloarchaea.com/resources/halohandbook/> version 7.2). All growth was at 45°C unless otherwise noted.

To introduce *trpA* at the *hdrB* locus of *H. volcanii*, we first inserted the *trpA* gene into the plasmid pTA160, originally designed to delete *hdrB*. The *trpA* gene, under the ferredoxin (*fdx*) promoter of *H. salinarium*, was amplified using primers AP389 (aaagctagcgcctcggtacccgggggatcc) and AP390 (tttgctagccgttatgtgcgttcggat), from pTA298. Using *Nhe*I, the PCR product was inserted into pTA160 between the *hdrB* flanking regions. Transformation of *H. volcanii* was carried out using the PEG method as described in (Cline et al. 1989).

Prior to hybridization, each culture was grown to an OD₆₀₀ of 1–1.1, and 2 ml samples were taken from both strains and applied to a 0.2 mm filter connected to a vacuum to eliminate excess media. The filter was then placed on a Petri dish containing a rich medium (HY medium + thymidine) for 48 hr at 42°C. The cells were washed and resuspended in Casamino broth, washed twice more in the same media, and plated on selective media.

Sequence library construction

RNA was isolated using EZ-RNA Total RNA Isolation Kit (Biological Industries Cat.# 20-400). DNA purification was done using the spooling method as described (<http://www.haloarchaea.com/resources/halohandbook/> version 7.2).

RNA-seq and DNA-seq libraries were prepared using Illumina TruSeq kits, following manufacturer protocols. All libraries were sequenced as multiplexed samples in one lane of an Illumina HiSeq 2000.

Genome annotation and read mapping

We obtained the genome assemblies and annotations for *H. volcanii* (strain DS2) and *H. mediterranei* (strain ATCC 33500) from NCBI RefSeq (accession numbers: GCF_000025685.1 and GCF_000337295.1, respectively). In order to determine which bases in each genome would be unambiguously mappable in the hybrids, in each parental genome, we employed a sliding window of 75 bp (our mapping read length; see below) and a step of one bp to create simulated NGS reads. These reads were mapped to a reference consisting of both parent's genomes using Bowtie 0.12.8, with default parameters, retaining only uniquely mapping reads. Any base overlapped by reads that could not be mapped uniquely were masked from further analysis (corresponding to 3.9% and 1.3% of the *H. volcanii* and *H. mediterranei* genomes, respectively.)

We identified orthologous genes between the two species using the RoundUp database (DeLuca et al. 2012). Genes were then grouped into operons based on the MicrobesOnline operon predictions in *H. volcanii* (Price et al. 2005; <http://meta.microbesonline.org/operons/gnc309800.html>). Corresponding *H. mediterranei*

operons were inferred from the presence of co-linearity of orthologs between the parental species.

All DNA-seq and RNA-seq reads were trimmed to 75 bp in length and mapped to a reference consisting of the concatenation of both parental genomes using Bowtie, version 0.12.8, with default parameters and retaining only uniquely mapping reads. DNA-seq RPKM was calculated using the number of unambiguously mappable bases as the gene length.

Detecting significant ASE

We determined base-level coverage of CDSs of both species for all uniquely mappable positions for both hybrid replicates for main chromosome located orthologs with at least 100 reads mapping among both alleles in both biological replicates (1,320 orthologs). As the DNA-seq data indicated that parental chromosomal abundance was not necessarily equal in both replicates, base-level coverages were linearly scaled such that the total coverage was identical for each species' main chromosome. The RNA-seq RPKMs were calculated as the base level coverage / (the number of uniquely mappable bases \times the total base level coverage for all orthologs \times the mapped read length [75 bp]). We applied the resampling test of Bullard et al. (2010): the base-level read coverage of each parental allele was resampled with replacement 10,000 times, under two conditions: 1) using the *H. volcanii* marginal nucleotide frequencies ($\delta_v = \delta_v[A], \delta_v[C], \delta_v[G], \delta_v[T]$) and the *H. volcanii* length, L_v , and 2) using the *H. mediterranei* marginal nucleotide frequencies δ_m and the *H. mediterranei* length, L_m . A started \log_2 ratio (total base level coverage from $\delta_v, L_v + 1$ / total base level coverage from $\delta_v, L_v + 1$), denoted as $\log_2(\delta_v^{Hv+1} / \delta_m^{Hm+1})$, was calculated from each resampling, and represents the variation in $\log_2(\delta_v^{Hv+1} / \delta_m^{Hm+1})$ *cis*-ratios expected between alleles due only to differential base frequencies and length. Both null

distributions (one per allele) were compared against the observed started $\log_2(\frac{H_{v+1}}{H_{m+1}})$ *cis*-ratio in order to obtain a two-tailed p value based on how often the observed ratio was outside of the bounds of the null distribution. In cases where both replicates agreed in the direction of parental bias, the least significant p-value among the four comparisons (both alleles in each replicate) was retained as a measure of the significance of differential expression. All p-values were adjusted such that we retained only those comparisons significant at an FDR of 5% for further analysis (Benjamini and Hochberg 1995).

To determine whether ASE measurements between genes within predicted operons were more similar those outside of operons, we performed 10,000 random samples of two categories of pairs of adjacent genes: either within predicted operons or outside of any predicted operon. For each sampled pair of genes we calculated the difference in the absolute values of \log_2 (ASE ratios). Finally, we asked whether the distribution of these differences from genes sampled within operons was significantly lower than that sampled outside of operons.

All statistical analyses were performed using R version 3.1.3 (R Core Team 2015). Kruskal-Wallis tests were performed using 10,000 permutations of the data as implemented in the ‘coin’ package (Hothorn et al. 2006).

Detecting selection on cis-regulatory divergence

Gene Ontology (GO) categories for *H. volcanii* genes were obtained from the EBI Quick-
GO database (accessed on 18 Feb. 2014; Bins et al. 2009). In the case of multi-gene operons, the operon was annotated as the union of the GO terms associated with its respective genes. For the purpose of interspecific comparisons, *H. mediterranei* orthologs were assigned to the same GO categories as *H. volcanii*.

Orthologs with significant *cis*-regulatory divergence at either level were divided into two categories based on the upregulating parental allele and ranked based on the magnitude of their absolute *cis* ratio (from largest to smallest). We searched for lineage-specific bias among GO biological process, GO molecular function, and GO cellular component. In order to detect lineage-specific bias within a gene set, we identified all functional categories containing at least 10 members in the set and determined whether significant bias existed in the direction of one or the other lineage using a χ^2 ‘goodness of fit’ test. Because many different categories were being tested, we determined the probability of observing a particular enrichment by permuting ortholog assignments and repeating the test 10,000 times, retaining the most significant p-value observed in each functional dataset. We obtained a permutation-based p-value by asking how often a χ^2 value of equal or greater significance would be observed in the permuted data. The test of bias was performed at two thresholds, using either the top 50% most biased orthologs, or analyzing all biased orthologs.

Growth in low phosphate

The low phosphate media was Hv-Min medium (Allers et al. 2004), supplemented with potassium phosphate buffer (pH 7.5), the only phosphate source, to a final concentration of 0.1 mM phosphate. To compare the growth rates each strain was grown in low phosphate minimal broth medium at 42°C in shaking incubator for three days to reach $OD_{600} > 0.4$, then both strains diluted to be at the same OD (<0.15) to start the growth analysis. The growth curves were done using a Biotek ELX808IU-PC in 96-well plates at 42°C with continuous shaking, measuring OD_{595} every 30 minutes for 30 hours.

Acknowledgements

We would like to thank the Fraser Lab for helpful discussions and advice, and R. Schreiber and S. Robinson for technical assistance. This work was supported by NIH grant 1R01GM097171-01A1 and ISF grant 535/15.

References

- Allers T, Ngo HP, Mevarech M, Lloyd RG. 2004. Development of additional selectable markers for the halophilic archaeon *Haloferax volcanii* based on the leuB and trpA genes. Appl Environ Microbiol. 70: 943-53.
- Artieri CG, Fraser HB. 2014. Evolution at two levels of gene expression in yeast. Genome Res. 24:411-421.
- Awise JC, Duvall SW. 1977. Allelic expression and genetic distance in hybrid macaque monkeys. J Hered. 68:23-30.
- Babak T, DeVeale B, Tsang EK, Zhou Y, Li X, Smith KS, Kukurba KR, Zhang R, Li JB, van der Kooy D, Montgomery SB, Fraser HB. 2015. Genetic conflict reflected in tissue-specific maps of genomic imprinting in human and mouse. Nat Genet 47: 544-9.
- Bartolomei MS, Zemel S, Tilghman SM. 1991. Parental imprinting of the mouse H19 gene. Nature 351:153-5.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B. 57:289–300.

- Binns D, Dimmer E, Huntley R, Barrell D, O'Donovan C, Apweiler R. 2009. QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics*. 25:3045-3046.
- Breuert S, Allers T, Spohn G, Soppa J. 2006. Regulated polyploidy in halophilic archaea. *PLoS One*. 1:e92.
- Bullard JH, Mostovoy Y, Dudoit S, Brem RB. 2010. Polygenic and directional regulatory evolution across pathways in *Saccharomyces*. *Proc Natl Acad Sci U S A*. 107:5058-5063.
- Cline SW, Lam WL, Charlebois RL, Schalkwyk LC, Doolittle WF. 1989. Transformation methods for halophilic archaeobacteria. *Can J Microbiol*. 35: 148-52.
- DeLuca TF, Cui J, Jung JY, St Gabriel KC, Wall DP. 2012. Roundup 2.0: enabling comparative genomics for over 1800 genomes. *Bioinformatics*. 28:715-716.
- Dickinson WJ, Rowan RG, Brennan MD. 1984. Regulatory gene evolution: adaptive differences in expression of alcohol dehydrogenase in *Drosophila melanogaster* and *Drosophila simulans*. *Heredity* 52: 215-25.
- Fraser HB 2011. Genome-wide approaches to the study of adaptive gene expression evolution. *Bioessays* 33: 469-77.
- Fraser HB. 2013. Gene expression drives local adaptation in humans. *Genome Res*. 23: 1089-96.

Fraser HB, Moses A, Schadt EE. 2010. Evidence for widespread adaptive evolution of gene expression in budding yeast. PNAS 107: 2977-82.

Fraser HB, Levy S, Chavan A, Shah HB, Perez JC, Zhou Y, Siegal ML, Sinha H. 2012. Polygenic cis-regulatory adaptation in the evolution of yeast pathogenicity. Genome Res 22:1930-9.

Fraser HB, Babak T, Tsang J, Zhou Y, Zhang B, Mehrabian M, Schadt EE. 2011. Systematic detection of polygenic cis-regulatory evolution. PLoS Genet 7:e1002023.

Hawkins M, Malla S, Blythe MJ, Nieduszynski CA, Allers T. 2013. Accelerated growth in the absence of DNA replication origins. Nature. 503:544-547.

H. Hitzeroth, J. Klose, S. Ohno, and U. Wolf. 1968. Asynchronous Activation of Parental Alleles at the Tissue-Specific Gene Loci Observed on Hybrid Trout During Early Development. Biochemical Genetics 1: 287-300.

Hothorn T, Hornik K, van de Wiel MA, Zeileis A. 2006. A Lego System for Conditional Inference. The American Statistician 60: 257-263.

House MA, Griswold CK, Lukens LN. 2014. Evidence for selection on gene expression in cultivated rice (*Oryza sativa*). Mol Biol Evol 31:1514-25.

Lange C, Zerulla K, Breuert S, Soppa J. 2011. Gene conversion results in the equalization of genome copies in the polyploid haloarchaeon *Haloferax volcanii*. Mol Microbiol. 80: 666-77.

Lazzari P, Solidoro C, Salon S, Bolzon G. 2016. Spatial variability of phosphate and nitrate in the Mediterranean Sea: A modeling approach. Deep Sea Research Part I: Oceanographic Research Papers 108: 39-52.

López-García P, St Jean A, Amils R, Charlebois RL. 1995. Genomic stability in the archaeae *Haloferax volcanii* and *Haloferax mediterranei*. J Bacteriol 177: 1405-8.

Mullakhanbhai MF, Larsen H. 1975. *Halobacterium volcanii* spec. nov., a Dead Sea halobacterium with a moderate salt requirement. Arch Microbiol 104: 207-14.

Naor A, Lapierre P, Mevarech M, Papke RT, Gophna U. 2012. Low species barriers in halophilic archaea and the formation of recombinant hybrids. Curr Biol 22: 1444-8.

Nissenbaum A, Stiller M, Nishri A. 1990. Nutrients in pore waters from Dead Sea sediments. Hydrobiologia 197: 83-89.

Ortenberg, R., Tchelet, R., & Mevarech, M. 1999. A model for the genetic exchange system of the extremely halophilic archaeon *Haloferax volcanii*. In A. Oren (Ed.), Microbiology and biogeochemistry of hypersaline environments (pp. 331–338). Boca Raton: CRC Press.

Pastinen T. 2010. Genome-wide allele-specific analysis: insights into regulatory variation. *Nat Rev Genet.* 11:533-8.

Price MN, Huang KH, Alm EJ, Arkin AP. 2005. A novel method for accurate operon predictions in all sequenced prokaryotes. *Nucleic Acids Res.* 33:880-892.

R Core Team. 2015. R: A language and environment for statistical computing.

R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

Rodriguez-Valera F, Juez G, Kushner DJ. 1983. *Halobacterium mediterranei* spec, nov., a New Carbohydrate-Utilizing Extreme Halophile. *Syst Appl Microbiol.* 4:369-81.

Rosenshine I, Tchelet R, Mevarech M. 1989. The mechanism of DNA transfer in the mating system of an archaeobacterium. *Science* 245:1387-9.

Stroud A, Liddell S, Allers T. 2012. Genetic and Biochemical Identification of a Novel Single-Stranded DNA-Binding Complex in *Haloferax volcanii*. *Front Microbiol* 3:224.

Wang X, Clark AG. 2014. Using next-generation RNA sequencing to identify imprinted genes. *Heredity* 113:156-66.

Wang X, Soloway PD, Clark AG. 2010. Paternally biased X inactivation in mouse neonatal brain. *Genome Biol.* 11:R79.

Wittkopp PJ, Haerum BK, Clark AG. 2004. Evolutionary changes in cis and trans gene regulation. *Nature* 430: 85-8.

Wittkopp PJ, Kalay G. 2011. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat Rev Genet.* 13:59-69.

Wright DA and Moyer FH. 1966. Parental Influences on Lactate Dehydrogenase in the Early Development of Hybrid Frogs in the Genus *Rana*. *J. Exp. Zool* 163: 215-230.

Figure Legends

Figure 1. Generation of stable *H. volcanii* - *H. mediterranei* hybrids. **A.** The genomic organization of the selectable markers involved in the study. **B.** Generation of a stable hybrid. H133 was transformed with pTA160 *trpA*, and upon selection on media lacking thymidine the *trpA* marker was integrated in the *hdrB* locus, generating UG241. UG241 was mated with WR646, which are autotrophs for thymidine and tryptophan, respectively. The mated colonies were selected on a media lacking thymidine and tryptophan. Small circles indicate the plasmids and the rectangle represents the chromosome.

Figure 2. Regulatory divergence between archaeal hybrids is revealed by ASE analysis. **(A)** Approximately equal numbers of orthologs show significant allelic bias favoring either the *H. mediterranei* (453, red) or *H. volcanii* allele (476, blue). RPKMs plotted in this figure are the mean of the two biological replicates after normalization (see Methods). med, mediterranei; vol, volcanii. **(B)** Pairwise comparisons of adjacent genes within predicted operons show significantly more similar ASE than independently transcribed adjacent genes (Kruskal-Wallis rank sum test, $p < 10^{-15}$ in both cases). Op, adjacent genes within predicted operons; Non-Op, adjacent genes outside of predicted operons.

Figure 3. Detection of lineage-specific selection and differential fitness in low phosphate conditions. **(A)** For a set of 19 genes, the expected number with ASE with higher expression from the *H. mediterranei* alleles is plotted. The kinase gene set had 15/19 genes favoring the *H.*

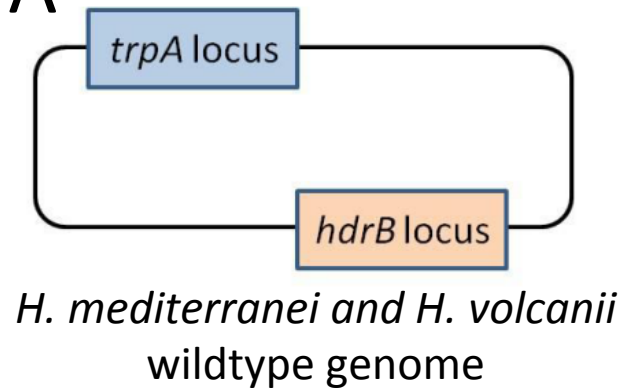
mediterranei alleles (red arrow), in both biological replicates. **(B)** *H. mediterranei* grows robustly in 0.1 mM phosphate, whereas *H. volcanii* does not. See also Supplemental Figure 3.

Supplemental Figure 1. DNA-seq data reveals gene amplifications in both parental chromosomes, as well as hybrid specific variation in plasmid copy number. The *H. volcanii* chromosomes of both hybrid cultures as well as the *H. mediterranei* chromosome of replicate 1 revealed segmental amplifications spanning one or multiple genes (indicated in grey; see supplementary methods). Furthermore, the ratio of copy number among plasmids and main chromosomes varied between the two hybrids. Note that two-fold copy number variation is expected along chromosomes and plasmids as cultures were collected during active DNA replication during log-phase growth (Hawkins et al. 2013). med, *mediterranei*; vol, *volcanii*.

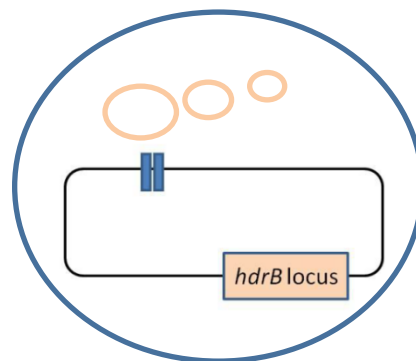
Supplemental Figure 2. ASE measurements are concordant between biological replicates. Shown are the (A) ASE values comparing each species' alleles across replicates; (B) ASE values comparing orthologs in each of the two biological replicates; and (C) as well as the inter-replicate correlation in ASE log-ratios. med, *mediterranei*; vol, *volcanii*.

Supplemental Figure 3. Biological replicate of growth curves in 0.1 mM phosphate.

A

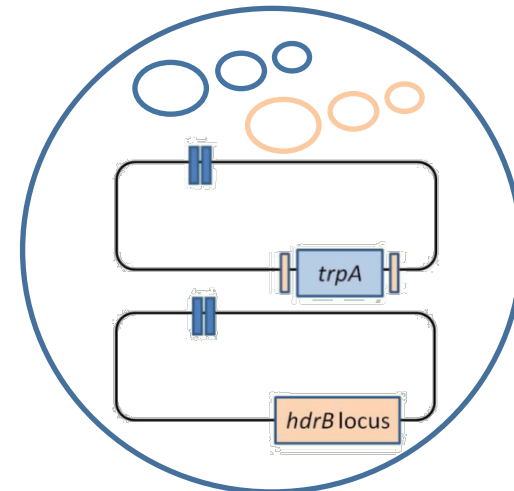


WR646: *H. mediterranei*
 $\Delta trpA$

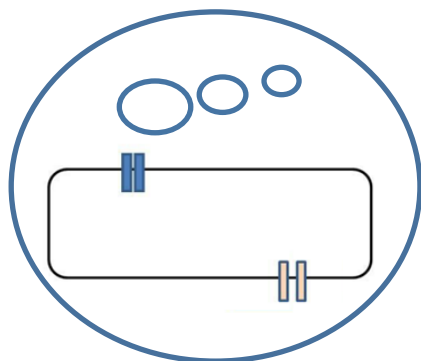


+

H. mediterranei/
H. volcanii hybrid

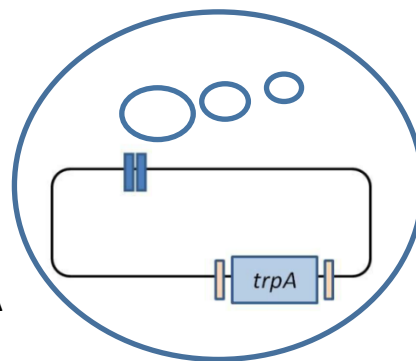


B



H133: *H. volcanii*
 $\Delta trpA \Delta hdrB$

Transformation
with pTA160-*trpA*



UG241: *H. volcanii*
 $\Delta trpA \Delta hdrB::trpA$

Fig. 2

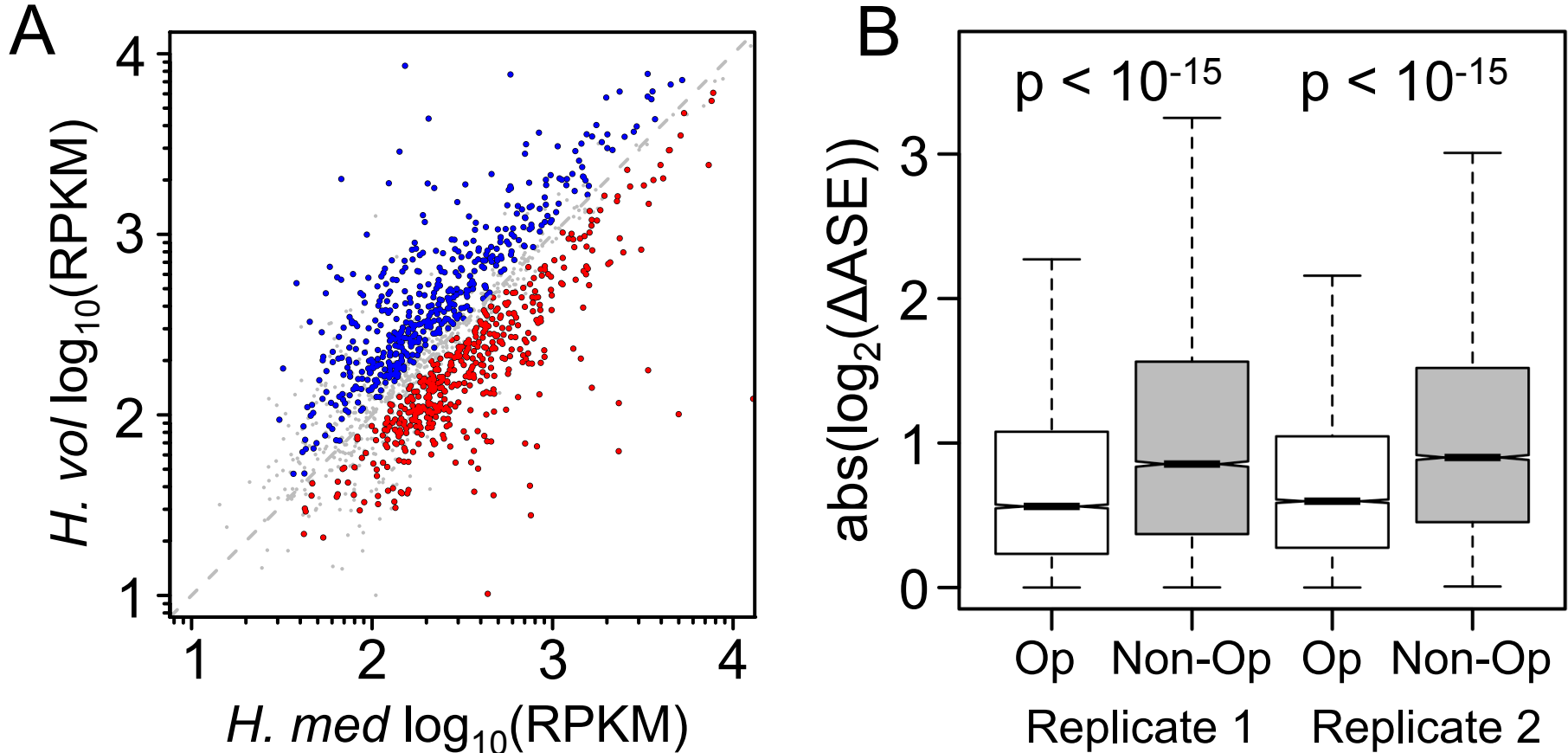
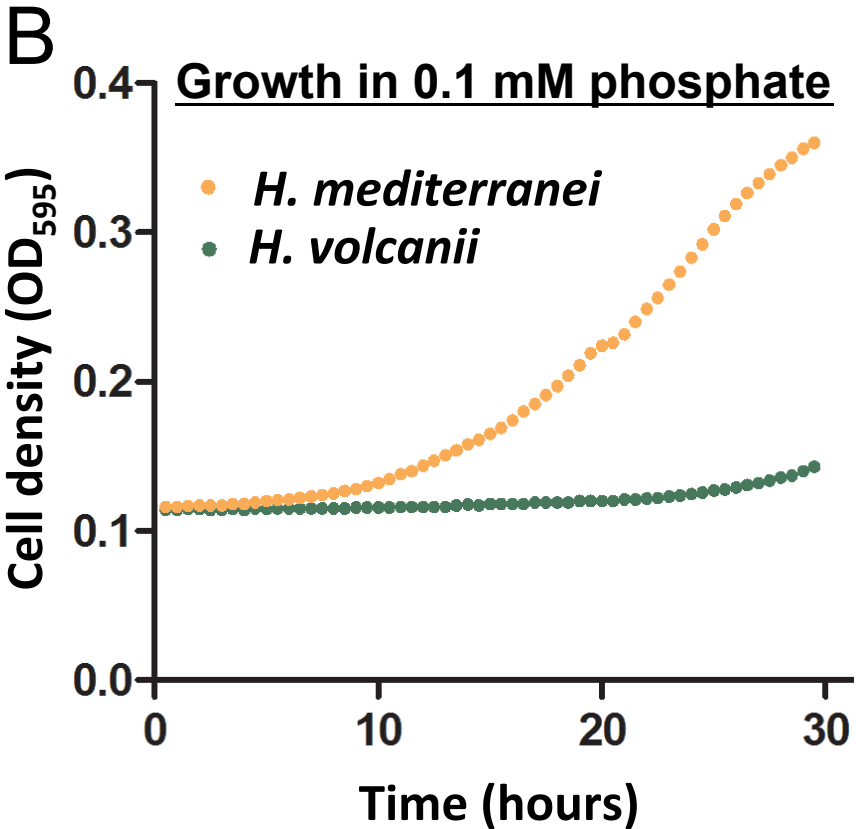
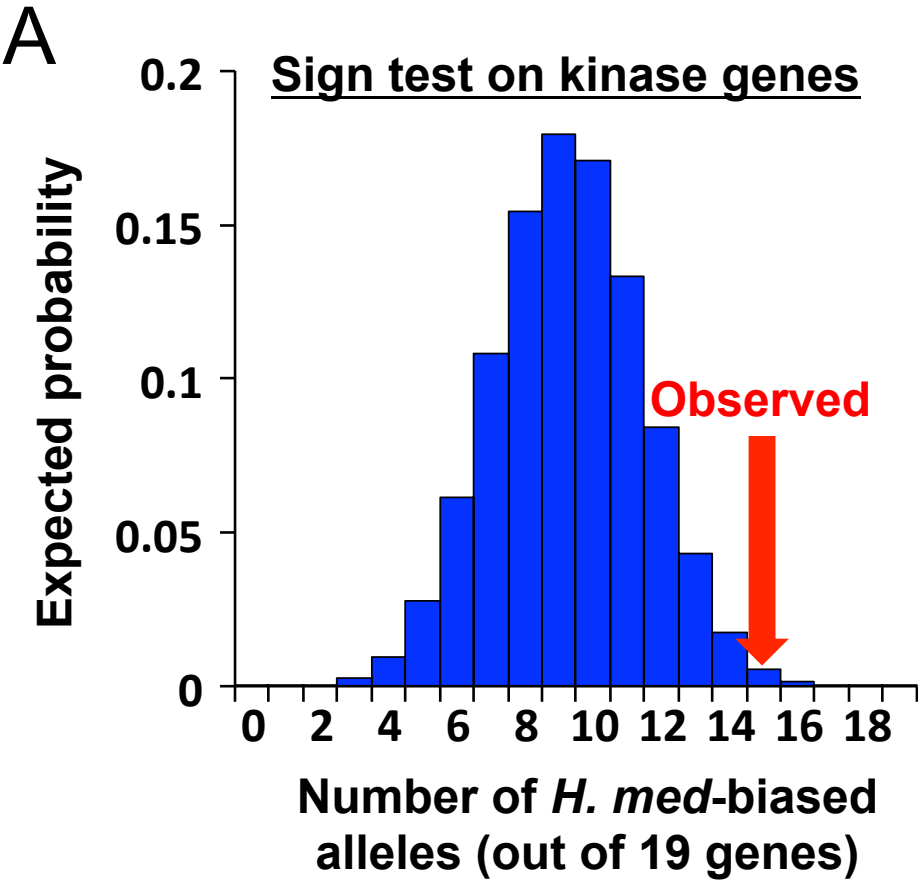
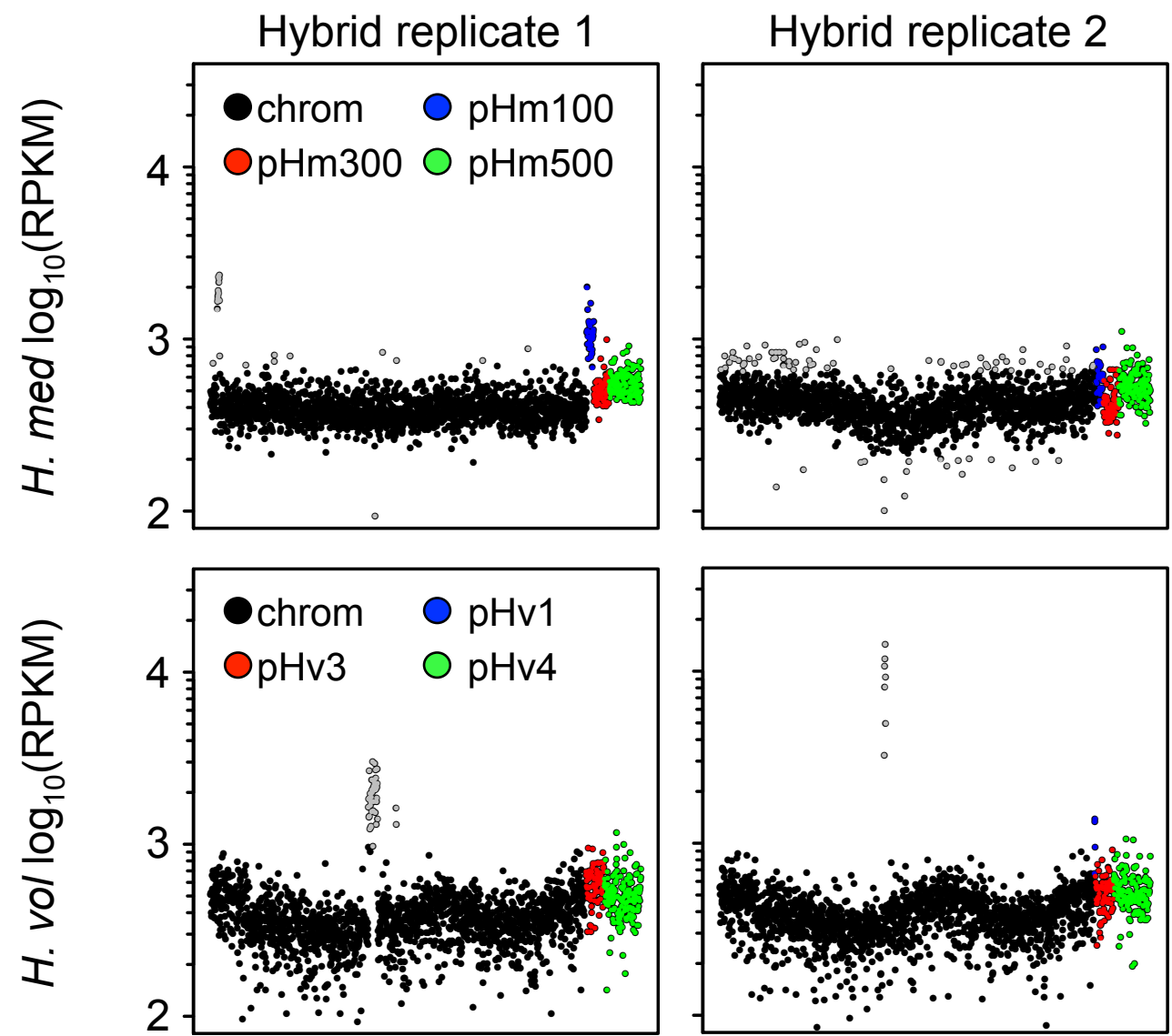


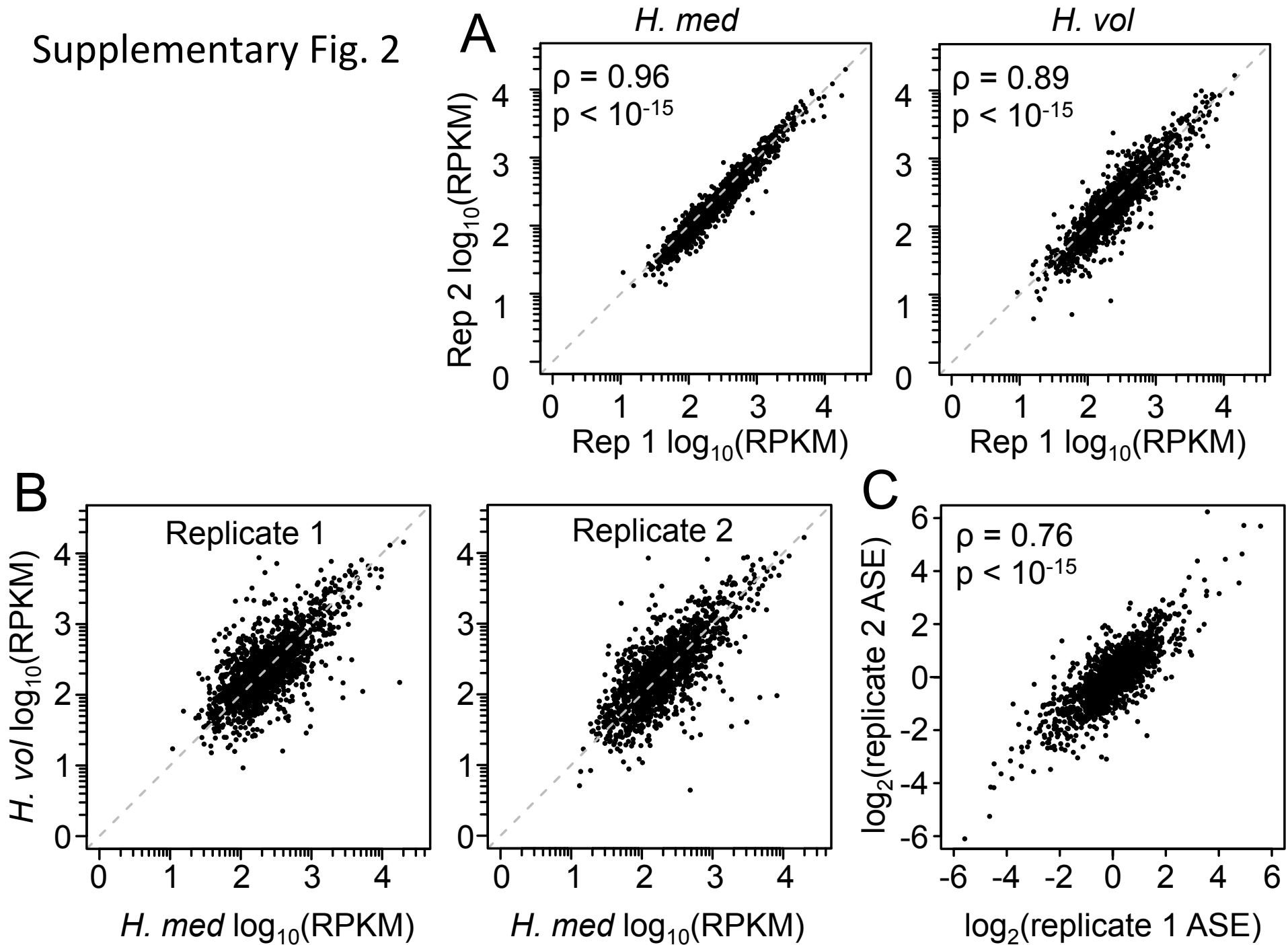
Fig. 3



Supplementary Fig. 1



Supplementary Fig. 2



Supplementary Fig. 3

