Sympatric speciation revealed by genome-wide divergence in the blind mole rat *Spalax*

Kexin Li[a,b,1], Wei Hong[a,1], Hengwu Jiao[a], Guo-Dong Wang[c], Karl A. Rodriguez[d,e], Rochelle Buffenstein[d,e,f], Yang Zhao[b], Eviatar Nevo[b,2], and Huabin Zhao[a,2]

[a]Department of Ecology, College of Life Sciences, Wuhan University, Wuhan 430072, China; [b]Institute of Evolution, University of Haifa, Haifa 3498838, Israel; [c]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China; [d]Sam and Anne Barshop Center for Longevity and Aging Studies, University of Texas Health Science Center at San Antonio, San Antonio, TX 78229; [e]Department of Physiology, University of Texas Health Science Center at San Antonio, San Antonio, TX 78229; and [f]Calico, South San Francisco, CA 94080

Contributed by Eviatar Nevo, July 29, 2015 (sent for review July 6, 2015; reviewed by Francisco J. Ayala, Sergey Gavrilets, and Alan Robert Templeton)

Sympatric speciation (SS), i.e., speciation within a freely breeding population or in contiguous populations, was first proposed by Darwin [Darwin C (1859) *On the Origins of Species by Means of Natural Selection*] and is still controversial despite theoretical support [Gavrilets S (2004) *Fitness Landscapes and the Origin of Species (MPB-41)*] and mounting empirical evidence. Speciation of subterranean mammals generally, including the genus *Spalax*, was considered hitherto allopatric, whereby new species arise primarily through geographic isolation. Here we show in *Spalax* a case of genome-wide divergence analysis in mammals, demonstrating that SS in continuous populations, with gene flow, encompasses multiple widespread genomic adaptive complexes, associated with the sharply divergent ecologies. The two abutting soil populations of *S. galili* in northern Israel habituate the ancestral Senonian chalk population and abutting derivative Plio-Pleistocene basalt population. Population divergence originated ∼0.2–0.4 Mya based on both nuclear and mitochondrial genome analyses. Population structure analysis displayed two distinctly divergent clusters of chalk and basalt populations. Natural selection has acted on 300+ genes across the genome, diverging *Spalax* chalk and basalt soil populations. Gene ontology enrichment analysis highlights strong but differential soil population adaptive complexes: in basalt, sensory perception, musculature, metabolism, and energetics, and in chalk, nutrition and neurogenetics are outstanding. Population differentiation of chemoreceptor genes suggests intersoil population's mate and habitat choice substantiating SS. Importantly, distinctions in protein degradation may also contribute to SS. Natural selection and natural genetic engineering [Shapiro JA (2011) *Evolution: A View From the 21st Century*] overrule gene flow, evolving divergent ecological adaptive complexes. Sharp ecological divergences abound in nature; therefore, SS appears to be an important mode of speciation as first envisaged by Darwin [Darwin C (1859) *On the Origins of Species by Means of Natural Selection*].

sympatric speciation | population genetics | genome divergence | ecological adaptation | natural selection

Despite more than a century since first proposed by Darwin (1), the concept of sympatric speciation (SS) as a major mode of speciation, i.e., formation of new species within a freely breeding population with ongoing gene flow, is still highly controversial and evaluated both skeptically and critically (2). Interestingly, recent empirical studies (*SI Appendix*) and theoretical assessments (3) support SS. Claims of SS must demonstrate species sympatry, sister relationship, reproductive isolation, and that an earlier allopatric phase is highly unlikely (2). We recently described two studies of SS in two evolutionarily divergent mammals, the blind subterranean mole rat *Spalax galili* (4) (Fig. 1 *A*–*C*) and the spiny mice, *Acomys*, at "Evolution Canyon" (EC), Mount Carmel, Israel (5). Moreover, the Evolution Canyon microsite in Israel has been suggested as a cradle for SS across life, based on incipient SS of five distant taxons: bacteria, wild barley, fruit flies, beetles, and spiny mice (6). Importantly, only a few studies to date have investigated whole genome evolution in an attempt to uncover genome architectural changes during SS (7). Here we show that SS in *S. galili* encompasses extensive adaptive complexes across the genome associated with the sharply abutting and divergent chalk and basalt ecologies where SS took place, i.e., in sympatry and not in an earlier allopatry (4).

Results

Population Sequencing and Variation Calling. Five individuals of the blind mole rat (*S. galili*) from the chalk rendzina soil and six from the abutting basalt soil (Fig. 1 *A*–*C*) were collected for genome sequencing. Of the 11 animals, the generated data for each individual, which had a genome size of ∼3G bp (8), ranged from 19.6 to 30.8 Gb, corresponding to sequencing depths of 6.36×–10× (Table S1). A total of 14,539,199 SNPs were identified, with 3,717,338 and 3,361,317 SNPs unique to the basalt and chalk populations, respectively (Fig. S1). We validated our SNP calling strategy with traditional Sanger sequencing technology and found that the genome-wide false-positive rate is less than 6% and the false-negative rate is less than 13% (*SI Materials and Methods*).

Significance

Sympatric speciation is still highly controversial. Here we demonstrate, based on genome-wide divergence analysis, that sympatric speciation in the blind subterranean rodent *Spalax galili* encompasses multiple and widespread genomic adaptive complexes associated with the sharply divergent and abutting basalt and chalk soil populations. Gene ontology enrichment analysis highlights sensory perception, musculature, metabolism, and energetics in basalt against neurogenetics and nutrition in chalk. Population divergence of chemoreceptor genes suggests the operation of mate and habitat choices, substantiating sympatric speciation. Natural selection and natural genetic engineering overrule gene flow, evolving divergent ecological adaptive complexes. Sympatric speciation may be a common speciation mode, as envisaged by Darwin, due to the abundance of sharp divergent geological, edaphic, climatic, and biotic ecologies in nature.

Author contributions: K.L., R.B., E.N., and H.Z. designed research; K.L., W.H., H.J., K.A.R., and Y.Z. performed research; W.H., H.J., G.-D.W., E.N., and H.Z. analyzed data; and K.L., W.H., H.J., K.A.R., E.N., and H.Z. wrote the paper.

Reviewers: F.J.A., University of California, Irvine; S.G., University of Tennessee; and A.R.T., Washington University at St. Louis.

The authors declare no conflict of interest.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession nos. SRP058797 and KT009027–KT012480).

[1]K.L. and W.H. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: nevo@research.haifa.ac.il or huabinzhao@whu.edu.cn.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1514896112/-/DCSupplemental.

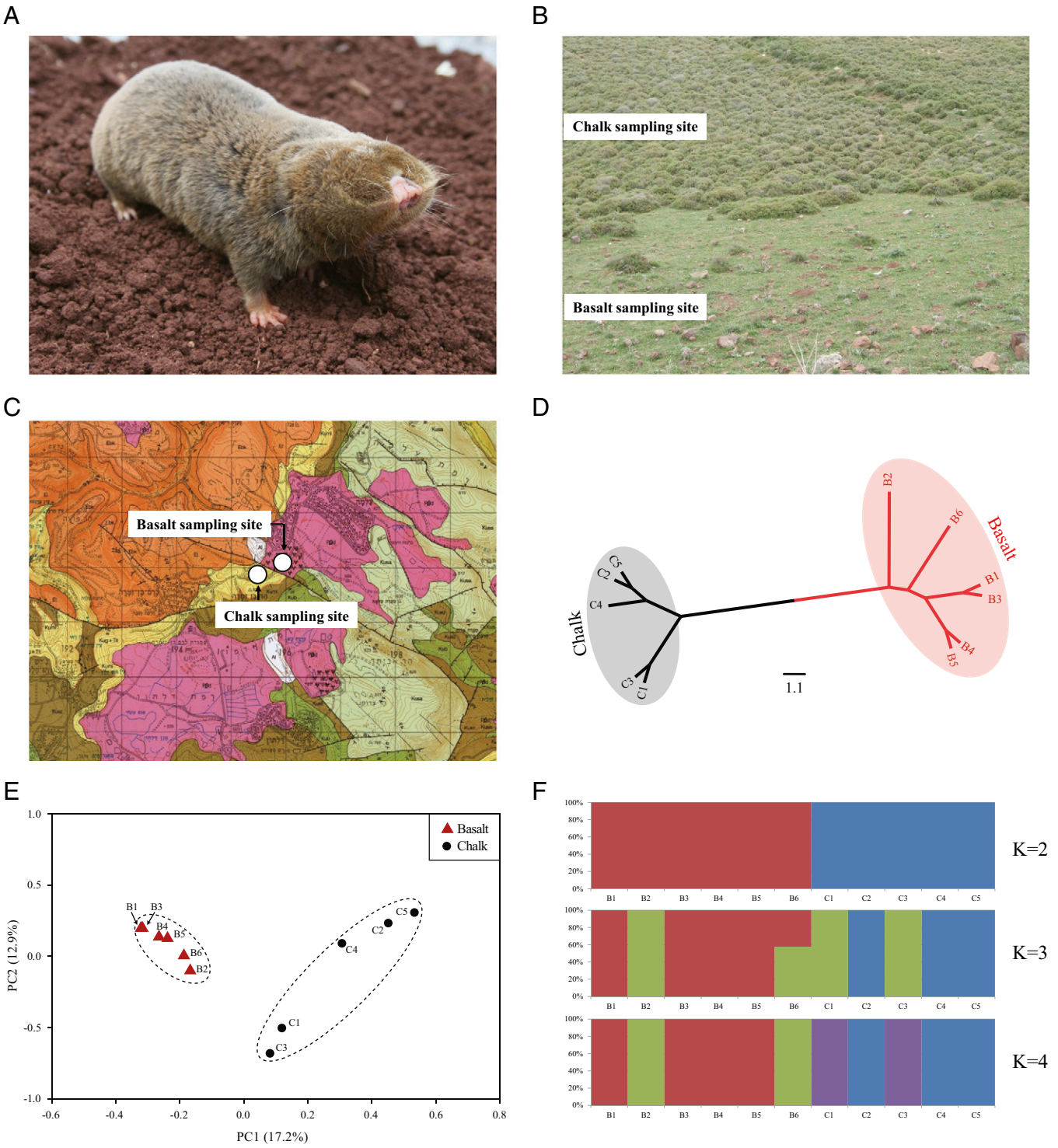

Fig. 1. Study subject, ecological, and genomic divergence in sympatric speciation. (*A*) The blind mole rat, *S. galili*. (*B*) Vegetation formation exposed with mounds in basalt and covered by dense bushes of *Sarcoptherium spinosum* in chalk. (*C*) Geological map. Chalk is in yellow and basalt is in pink, separated by a geological fault. (*D*) NJ tree based on all of the SNPs. Red, basalt individuals; black, chalk individuals. (*E*) PCA of two *Spalax* soil populations. Red triangles, basalt individuals; black circles, chalk individuals. (*F*) Population genetic structure of two *Spalax* soil populations when K = 2, 3, and 4.

Genetic Diversity and Structure of the Two Populations. Whole genome genetic diversity, estimated by Watterson's θ for each individual, was significantly higher in chalk (mean $\theta = 1.13 \times 10^{-3}$) than in basalt ($1.09 \times 10^{-3}$; $P < 2.2 \times 10^{-16}$, Mann–Whitney U test). Two distinct soil population clusters, basalt and chalk, were identified by the neighbor-joining (NJ) method (Fig. 1*D*) and by principal component analysis (PCA) (Fig. 1*E*) based on all SNPs. In PCA, the first and second component explained 17.2% and 12.9% of the genetic differences, respectively. A maximum likelihood analysis of population structure was undertaken based on K = 2, 3, and 4 (Fig. 1*F*). With K = 2, the basalt population is completely separated from the chalk population. In clustering by

K = 3, both populations were divided into further subgroups but with the same genetic background (in green), notably, the individual B6 spans both chalk and basalt populations, possibly a recombinant individual. With K = 4, both populations were split into two subgroups, respectively (Fig. 1*F*).

Linkage Disequilibrium and Population Demography. Linkage disequilibrium (LD), measured by the correlation coefficient (r^2), decreased rapidly below 0.2 within 1,000 bp in both populations (Fig. 2*A*). Mean population recombination rate per kilobase ($\rho = 4Ne \times r$) in basalt was 1.515, significantly higher than in chalk of 1.125 ($P < 2.2 \times 10^{-16}$, Mann–Whitney *U* test). Estimated extant effective population size (*Ne*) for chalk and basalt was 72,380 and 116,800, respectively, and the inferred ancestral *Ne* was 84,570 (Fig. 2*B*). The migration rates of chalk to basalt and basalt to chalk were 1.788 and 5.809 per generation (Fig. 2*B*), respectively. The two *Spalax galili* soil populations were estimated to split ~0.2–0.4 Mya based on both whole genome and mitochondrial genome analyses (Fig. 2*B* and *SI Materials and Methods*).

Population Genomic Divergence and Functional Enrichment. Putatively selected genes (PSGs) were identified by screening genomic regions that show low diversity (measured by Tajima' *D*) in one population but high divergence [measured by fixation index (F_{ST})] between the two populations. A total of 128 genes in the basalt population and 189 genes in the chalk population were identified with the strongest signature of positive selection (Tables S2 and S3). Gene ontology (GO) enrichment analysis revealed functions of musculature, energetics, metabolism, and sensory biology (Fig. 3) distinguished in basalt. In contrast, functions related to neurogenetics and nutrition were enriched in the chalk population (Fig. 3). Our genome-wide scans for positive selection highlight how, where, and why adaptive evolution has shaped genetic variations.

We sequenced 22 olfactory receptor (*OR*) genes (i.e., genes are numbered in Table 1) (Fig. S2), 20 bitter taste receptor (*Tas2r*) genes (Fig. S3), and 18 putatively neutral noncoding regions (*NCs*). Specifically, *ORs* refer to a group of olfactory receptor genes; *Tas2rs* refer to a group of type 2 taste receptor genes; *NCs* refer to a group of neutral noncoding regions. Pairwise F_{ST} estimates revealed that 10 *OR*s, 9 *Tas2r*s, and 1 neutral region are significantly differentiated between the chalk and basalt populations [$P < 0.05$ after false discovery rate (FDR) adjustment; Table 1]. The rate of significantly differentiated loci is statistically higher in *ORs* (10/22 = 45.5%) than in *NCs* (1/18 = 5.6%) ($P = 0.011$, two-tailed Fisher exact test), and the same is true for *Tas2r*s ($P = 0.009$).

Differences in Proteostasis of the Two Populations. Besides genomics, proteomics, particularly mechanisms regulating proteostasis, selectively drive soil population divergence. Mesic basalt population has a threefold significantly greater proteasome chymotrypsin-like activity in muscle and twofold higher trypsin-like and caspase-like activities than the chalk population (Fig. 4*A*) based on analyses of seven animals from each population. The basalt population displayed significantly higher α7 levels (Fig. 4*B*) than the chalk population, suggesting higher numbers of proteasomes in these tissues facilitating enhanced degradation of damaged or misfolded proteins through this protein degradation machinery. In contrast, the chalk population showed significantly higher levels of ATG7 and autophagic flux (LC3II/LC3I ratio) (Fig. 4*C*). This protein degradation profile suggested more of a reliance on autophagy in chalk population.

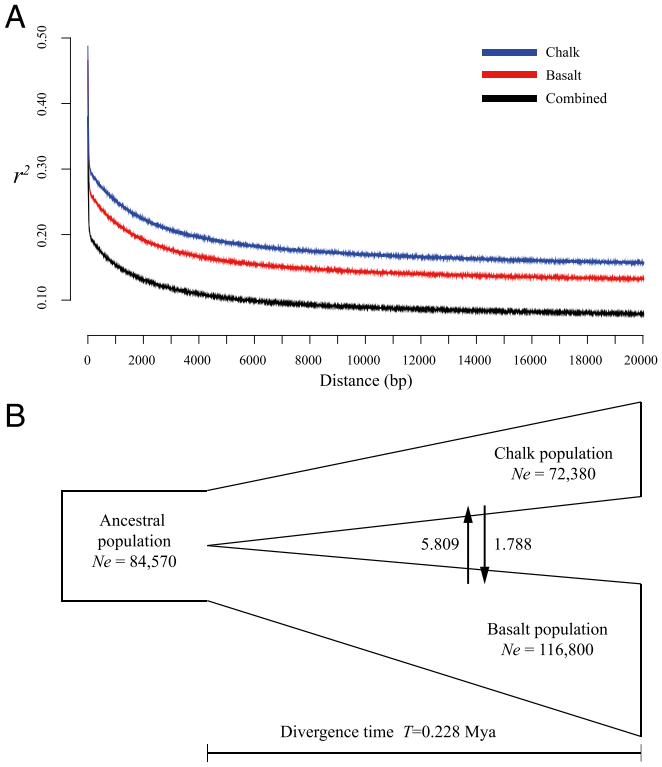

Fig. 2. LD and demographic structure. (*A*) LD decay of two continuous *S. galili* populations. *x* axis stands for physical distances (bp), whereas *y* axis stands for r^2. (*B*) Inferred demographic history for the two abutting soil populations (chalk vs. basalt) of *S. galili*. The extant and ancestral population sizes (*Ne*) of the chalk and basalt populations are indicated, and the migration rates between the two populations are provided. The divergence time (*T*) between two populations was inferred.

Discussion

Genetic Diversity, Population Structure, Speciation Genes, and Reproductive Isolation. Genetic diversity at the local chalk-basalt state parallels the regional patterns (9, 10). The basalt and chalk mole rat populations were grouped separately (Fig. 1 *D–F*), suggesting genome-wide divergence of the two abutting soil populations. Population structure analysis showed similar clustering as revealed by mtDNA (4). The B6 individual showed a genetic mixture of both the chalk and basalt populations, which may indicate some limited gene flow, as was shown earlier by a recombinant individual in the interface of chalk and basalt (4) and a mound row extending from basalt to chalk was also observed in the field. Remarkably, genetic diversity from whole genome analysis is consistent with earlier estimates by mtDNA (4) and AFLP (11), where genetic polymorphism was significantly higher in chalk than in basalt. This finding supports earlier evidence that genetic diversity is possibly associated with a more stressed environment (5, 6, 9, 10). We deduced, across the whole genome, positively selected speciation genes (SGs). We hypothesize that besides highlighting strong selection, additional factors could potentially contribute directly or indirectly to reproductive isolation (RI) in abutting *Spalax* soil populations associated with SS (Table S4). These factors include habitat selection, chemoreceptor divergence, and preliminary indications of mate choice.

LD and Population Demography. Rapid LD decay indicates high recombination rates or large effective population sizes (12). LD level is relatively lower in basalt than in the chalk population, possibly because of the larger basalt population size (4) (Fig. 2*B*) and the higher temperature stress in chalk as was shown experimentally in *Drosophila* (13). *Ne* of chalk is smaller than that of basalt (4) (Fig. 2*B*), probably due to lower food resources (14). The larger *Ne* of the basalt population (Fig. 2*B*) is similar to current population estimates (14). The recombination rate is significantly

EVOLUTION

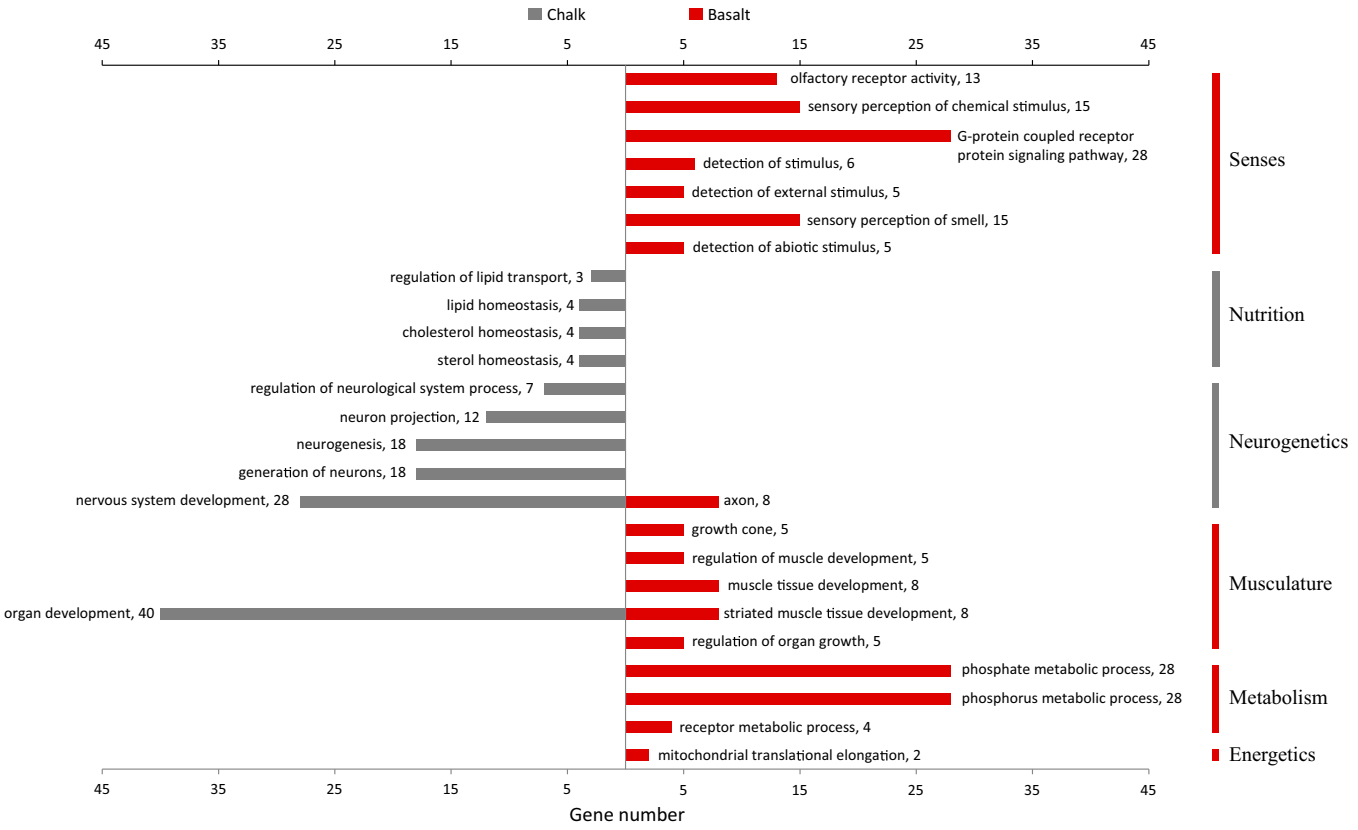

Fig. 3. GO enrichment analysis of putatively selected genes in chalk and basalt *S. galili* populations. Gene number is provided next to each GO term.

higher in basalt than that in chalk possibly due to adaptation to the novel basalt niche (15). The combined lower LD may suggest that LDs are soil specifically selected for the chalk and basalt separately rather than generally for both. There are more animals migrating from basalt to chalk populations possibly because of the higher population density in basalt.

Functional Adaptive Complexes. Functional enrichment analysis of GO terms revealed the remarkable divergence of biological differences between the abutting *Spalax* soil populations, indicating genome-wide adaptive complexes in each soil population resulting from its soil-unique, multiple-specific ecological stresses. Burrow digging, a major activity of mole rats, is harder

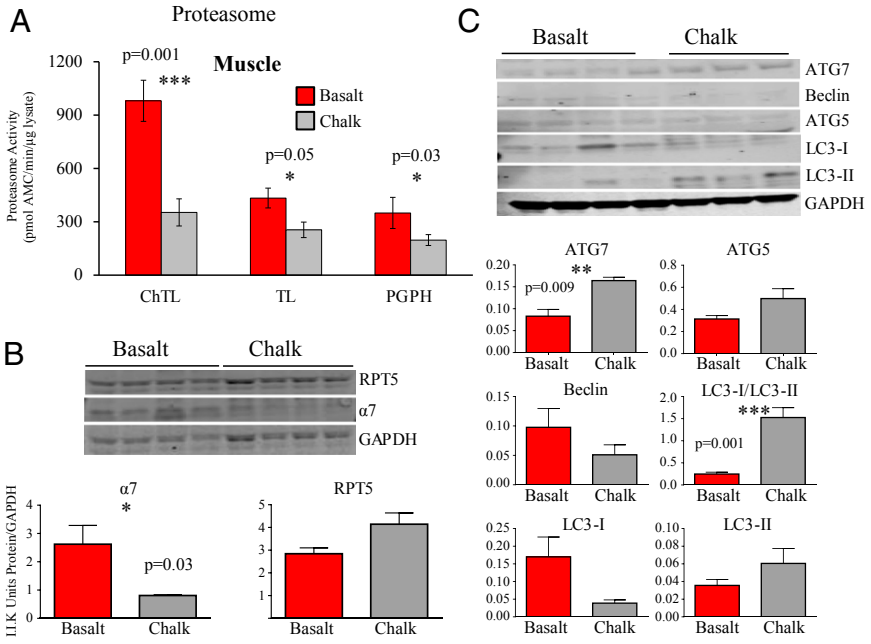

Fig. 4. Differences in the proteolytic machinery of *Spalax galili* basalt and chalk populations. (*A*) 20S proteasome activity measured by degradation of optimized peptides cleaved by the three active sites, chymotrypsin-like (ChTL), trypsin-like (TL), and post-glutamyl, peptide hydrolyzing (PGPH) or caspase-like, show higher levels of activity in the basalt population. (*B*) Higher levels of the constitutive 20S proteasome subunit α7 support the observation of the increase in activity in basalt population. (*C*) The chalk population protein degradation profile suggests more of a reliance on autophagy, with significantly higher levels of ATG7 and autophagic flux (LC3II/LC3I ratio).

Table 1. Pairwise F_{ST} statistics of olfactory receptor genes (*OR1-OR22*), bitter taste receptor genes (*Tas2r1-Tas2r20*), and putatively neutral noncoding regions (*NC1-NC18*) between basalt and chalk populations

Loci	F_{ST}	FDR
OR1	0.074	0.093
OR2	0.055	0.029
OR3	0.037	0.051
OR4	0.061	0.011
OR5	0.320	0.000
OR6	0.106	0.029
OR7	0.023	0.465
OR8	0.029	0.093
OR9	0.040	0.220
OR10	0.064	0.188
OR11	−0.003	0.663
OR12	0.000	0.220
OR13	−0.004	0.051
OR14	0.120	0.033
OR15	0.107	0.040
OR16	0.007	0.037
OR17	0.040	0.029
OR18	0.049	0.208
OR19	0.008	0.029
OR20	0.014	0.042
OR21	−0.006	0.155
OR22	0.050	0.169
Tas2r1	−0.025	0.752
Tas2r2	−0.014	0.354
Tas2r3	0.045	0.016
Tas2r4	0.023	0.297
Tas2r5	−0.018	0.604
Tas2r6	−0.001	0.072
Tas2r7	0.000	0.109
Tas2r8	0.000	0.752
Tas2r9	0.015	0.157
Tas2r10	0.166	0.026
Tas2r11	0.104	0.047
Tas2r12	−0.013	0.540
Tas2r13	0.052	0.035
Tas2r14	0.005	0.467
Tas2r15	0.130	0.023
Tas2r16	0.127	0.000
Tas2r17	0.320	0.000
Tas2r18	−0.004	0.516
Tas2r19	0.056	0.000
Tas2r20	0.080	0.016
NC1	0.014	0.458
NC2	−0.008	0.418
NC3	0.037	0.144
NC4	0.056	0.268
NC5	0.064	0.413
NC6	0.061	0.063
NC7	0.042	0.261
NC8	0.005	0.694
NC9	0.050	0.153
NC10	0.002	0.413
NC11	−0.035	0.939
NC12	0.064	0.261
NC13	0.030	0.153
NC14	0.021	0.018
NC15	0.018	0.371
NC16	0.017	0.153
NC17	−0.011	0.413
NC18	−0.009	0.458

Significant *P* values are underlined. The genes and noncoding regions were numbered in arbitrary order. FDR, false discovery rate; F_{ST} is the fixation index.

in compact basalt than in loose chalk (4), evidenced by several GO terms such as musculature, energetics, and sensory biology (Fig. 3). Accordingly, activity patterns (4) and mound density (14) are significantly higher in basalt than in chalk. Remarkably, the clay-rich basalt is harder and sticky, requiring more energy in digging. The enriched metabolic PSGs in basalt reflect the higher resting metabolic rate (RMR) in basalt than in chalk mole rats (4), as is true regionally in northern *S. galili* compared with southern *S. judaei* due to low food resources in the latter (16). Olfaction, developed in *Spalax* (17), is under stronger selection in basalt than in the chalk population, possibly due to higher food diversity and chemical resources in basalt. Likewise, soil moisture is higher in basalt than in chalk (14), increasing odor saturability in the soil and enhancing olfactory reception (18, 19). Finally, sensory perception of chemical stimuli may increase during the colonization of the chalk ancestors into the novel basalt niche (20) due to the higher level of selective pressure for mate choice.

This local pattern of neurogenetics and nutrition parallels the regional one of *S. judaei,* characterized by high genetic diversity (21), limited food resources, and relatively bigger brain size (22). Both locally (chalk population of *S. galili*) and regionally (*S. judaei*), food resources are limited; hence, larger territories and magnetic orientation from the nest to the periphery influence selection for advanced neurogenetics (22) (Fig. 3). Local and regional parallel patterns highlight the adaptive nature of the above traits.

Genetic variations in olfactory and taste receptors may underlie both mate and habitat choice in speciation (23). The divergence of both olfactory and taste receptor genes strongly suggest that olfaction and taste have undergone diversifying selection in population divergence and premating reproductive isolation. Thus, chemosensory habitat and mate preferences may substantiate SS in the blind mole rats.

Greater digging in harder basalt may manifest in greater muscle remodeling, a process dependent on the proteasome (24), leading to higher rates of proteasome mediated protein degradation and higher levels of constitutive proteasome components (α7 levels), as seen in the basalt new derivative species (Fig. 4*B*). High proteasome activity enhances the removal of damaged proteins and thereby improves the quality of the proteome and the maintenance of proteostasis. This may increase longevity (25, 26), and helps the cell to resist oxidative stress (25, 27). High levels of proteasome activity have also been seen in other underground-burrowing animals, such as the naked mole rat. Alternatively, compromised nutrition from low food resources as seen in chalk *Spalax*, which relies on lipid metabolism (Fig. 3), may induce autophagy rather than proteasome-mediated degradation for protein recycling (28) (Fig. 4*C*).

Conclusions and Prospects

During SS, natural selection and natural genetic engineering (29) are nonrandomly diverging 317 soil-biased genes, gene networks, and functional structures involved in the whole genome and proteome, coping with local ecologically, differentially, diverse biological stresses (Fig. 3 and Tables S2 and S3). Because the soil populations are abutting, animals can migrate in both directions (4) (Fig. 2*B*), yet philopatry or habitat choice restricts gene flow. Natural selection and natural genetic engineering overrule gene flow, as in Evolution Canyon (6), selecting for soil-specific adaptive complexes and generating adaptive ecological incipient SS. Unfolding the regulatory function of the repeatome, and epigenomics could highlight the SS procession in sharply divergent ecologies, which may occur frequently in nature (1) across sharply divergent geological, climatic, edaphic, and biotic abutting contrasts (6). Basalt populations have more SGs, which may involve genes explained by the niche release hypothesis (30). The *Spalax* example of SS supported by genome-wide adaptive

divergence is one of the few cases (*SI Appendix*) substantiating the occurrence and genetic patterns of SS in nature. Future studies could involve transcriptomes, repeatomes, metabolomics, editomes, and phenomics aspects of the ancestral chalk and derivative basalt populations.

Materials and Methods

All experiments on the blind mole rats were approved by the Ethics Committee of University of Haifa and Wuhan University, and conformed to the rules and guidelines on animal experimentation in Israel and China. Whole genome resequencing of two soil populations of *S. galili*, the ancestral chalk and derivative basalt populations, was performed. Genome-wide divergence between the two mole rat soil populations was estimated by PCA, neighbor-joining phylogenetic tree, and individual ancestry estimation based on the full maximum likelihood method. Genomic characterization of the two soil populations was revealed by differently enriched GO terms, protein proteostasis, olfactory, and bitter taste receptor gene analyses. Genetic diversity and LD of the two soil populations were compared and contrasted. Population demography, effective population size, and gene flow were estimated to assess the evolutionary divergence of sympatric speciation. Population divergent time was estimated by both mitochondrial and nuclear genome analyses. Full details of the materials and methods are described in *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank Professor Alan R. Templeton and Matěj Lövy for valuable comments and Shay Zur for his photograph. The Major Research Plan of the National Natural Science Foundation of China (91331115) and the Ancell-Teicher Research Foundation for Genetics and Molecular Evolution financially supported this work.

1. Darwin C (1859) *On the Origins of Species by Means of Natural Selection* (John Murray, London).
2. Coyne JA, Orr HA (2004) *Speciation* (Sinauer Associates, Sunderland, MA).
3. Gavrilets S (2004) *Fitness Landscapes and the Origin of Species (MPB-41)* (Princeton Univ Press, Princeton).
4. Hadid Y, et al. (2013) Possible incipient sympatric ecological speciation in blind mole rats (*Spalax*). *Proc Natl Acad Sci USA* 110(7):2587–2592.
5. Hadid Y, et al. (2014) Sympatric incipient speciation of spiny mice *Acomys* at "Evolution Canyon," Israel. *Proc Natl Acad Sci USA* 111(3):1043–1048.
6. Nevo E (2014) Evolution in action: Adaptation and incipient sympatric speciation with gene flow across life at "Evolution Canyon", Israel. *Isr J Ecol Evol* 60(2-4):85–98.
7. Michel AP, et al. (2010) Widespread genomic divergence during sympatric speciation. *Proc Natl Acad Sci USA* 107(21):9724–9729.
8. Fang X, et al. (2014) Genome-wide adaptive complexes to underground stresses in blind mole rats *Spalax*. *Nat Commun* 5:3966.
9. Nevo E (1999) *Mosaic Evolution of Subterranean Mammals: Regression, Progression, and Global Convergence* (Oxford University Press, Oxford, UK).
10. Nevo E (1998) Molecular evolution and ecological stress at global, regional and local scales: The Israeli perspective. *J Exp Zool* 282(1-2):95–119.
11. Polyakov A, Beharav A, Avivi A, Nevo E (2004) Mammalian microevolution in action: Adaptive edaphic genomic divergence in blind subterranean mole-rats. *Proc Biol Sci* 271(Suppl 4):S156–S159.
12. Reich DE, et al. (2001) Linkage disequilibrium in the human genome. *Nature* 411(6834): 199–204.
13. Franssen SU, Nolte V, Tobler R, Schlötterer C (2015) Patterns of linkage disequilibrium and long range hitchhiking in evolving experimental *Drosophila melanogaster* populations. *Mol Biol Evol* 32(2):495–509.
14. Lövy M, et al. (2015) Habitat and burrow system characteristics of the blind mole rat *Spalax galili* in an area of supposed sympatric speciation. *PLoS One* 10(7):e0133157.
15. Smukowski CS, Noor MA (2011) Recombination rate variation in closely related species. *Heredity (Edinb)* 107(6):496–508.
16. Nevo E, Shkolnik A (1974) Adaptive metabolic variation of chromosome forms in mole rats, *Spalax*. *Experientia* 30(7):724–726.
17. Heth G, et al. (1992) Differential olfactory perception of enantiomeric compounds by blind subterranean mole rats (*Spalax ehrenbergi*). *Experientia* 48(9):897–902.
18. Heth G, Todrank J, Nevo E (2000) Do *Spalax ehrenbergi* blind mole rats use food odours in searching for and selecting food? *Ethol Ecol Evol* 12(1):75–82.
19. Li M, et al. (2013) Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nat Genet* 45(12):1431–1438.
20. Schrader L, et al. (2014) Transposable element islands facilitate adaptation to novel environments in an invasive species. *Nat Commun* 5:5495.
21. Nevo E, Filippucci MG, Beiles A (1994) Genetic polymorphisms in subterranean mammals (*Spalax ehrenbergi* superspecies) in the near east revisited: Patterns and theory. *Heredity (Edinb)* 72(Pt 5):465–487.
22. Nevo E, Pirlot P, Beiles A (1988) Brain size diversity in adaptation and speciation of subterranean mole rats. *J Zoological Syst Evol Res* 26(6):467–479.
23. Smadja C, Butlin RK (2009) On the scent of speciation: The chemosensory system and its role in premating isolation. *Heredity* 102(1):77–97.
24. Rodriguez KA, Edrey YH, Osmulski P, Gaczynska M, Buffenstein R (2012) Altered composition of liver proteasome assemblies contributes to enhanced proteasome activity in the exceptionally long-lived naked mole-rat. *PLoS One* 7(5):e35890.
25. Rodriguez KA, et al. (2011) Walking the oxidative stress tightrope: A perspective from the naked mole-rat, the longest-living rodent. *Curr Pharm Des* 17(22):2290–2307.
26. Chondrogianni N, Georgila K, Kourtis N, Tavernarakis N, Gonos ES (2015) 20S proteasome activation promotes life span extension and resistance to proteotoxicity in *Caenorhabditis* elegans. *FASEB J* 29(2):611–622.
27. Pickering AM, Davies KJ (2012) Degradation of damaged proteins: The main function of the 20S proteasome. *Prog Mol Biol Transl Sci* 109:227–248.
28. Lin T-C, et al. (2012) Autophagy: Resetting glutamine-dependent metabolism and oxygen consumption. *Autophagy* 8(10):1477–1493.
29. Shapiro JA (2011) *Evolution: A View From the 21st Century* (Pearson Education, FT Press Science, Upper Saddle River, NJ).
30. Wu C-I, Ting C-T (2004) Genes and speciation. *Nat Rv Genet* 5(2):114–122.
31. Levitte D (2001) *Geological Map of Israel, 1: 50,000 Sheet 2–III, Zefat* (Geological Survey of Israel, Jerusalem).
32. Nevo E, Ivanitskaya EN, Beiles A (2001) *Adaptive Radiation of Blind Subterranean Mole Rats: Naming and Revisiting the Four Sibling Species of the Spalax ehrenbergi Superspecies in Israel: Spalax galili (2n=52), S. golani (2n=54), S. carmeli (2n=58), and S. judaei (2n=60)* (Backhuys Publishers, Leiden).
33. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760.
34. Li H, et al.; 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
35. Menozzi P, Piazza A, Cavalli-Sforza L (1978) Synthetic maps of human gene frequencies in Europeans. *Science* 201(4358):786–792.
36. Bryc K, et al. (2010) Genome-wide patterns of population structure and admixture in West Africans and African Americans. *Proc Natl Acad Sci USA* 107(2):786–791.
37. Liu Y, et al. (2013) Softwares and methods for estimating genetic ancestry in human populations. *Hum Genomics* 7(1):1.
38. Purcell S, et al. (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81(3):559–575.
39. Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: A tool for genome-wide complex trait analysis. *Am J Hum Genet* 88(1):76–82.
40. Tang H, Peng J, Wang P, Risch NJ (2005) Estimation of individual admixture: analytical and study design considerations. *Genet Epidemiol* 28(4):289–301.
41. Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4(4):406–425.
42. Browning SR, Browning BL (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet* 81(5):1084–1097.
43. Team RDC (2010) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna).
44. Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* 7(2):256–276.
45. Auton A, McVean G (2007) Recombination rate estimation in the presence of hotspots. *Genome Res* 17(8):1219–1227.
46. Gronau I, Hubisz MJ, Gulko B, Danko CG, Siepel A (2011) Bayesian inference of ancient human demography from individual genome sequences. *Nat Genet* 43(10): 1031–1034.
47. Rambaut A, Suchard M, Xie D, Drummond A (2014) Tracer v1. 6. Computer program and documentation distributed by the author. Available at beast.bio.ed.ac.uk/Tracer. Accessed July 27, 2014.
48. Gibbs RA, et al. (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 428(6982):493–521.
49. Adkins RM, Walton AH, Honeycutt RL (2003) Higher-level systematics of rodents and divergence time estimates based on two congruent nuclear genes. *Mol Phylogenet Evol* 26(3):409–420.
50. Shanas U, Heth G, Nevo E, Shalgi R, Terkel J (1995) Reproductive behaviour in the female blind mole rat (*Spalax ehrenbergi*). *J Zool (Lond)* 237(2):195–210.
51. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7(1):214.
52. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123(3):585–595.
53. Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38(6):1358–1370.
54. Dennis G, Jr, et al. (2003) DAVID: Database for annotation, visualization, and integrated discovery. *Genome Biol* 4(5):3.
55. Yarmolinsky DA, Zuker CS, Ryba NJ (2009) Common sense about taste: From mammals to insects. *Cell* 139(2):234–244.
56. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30(12):2725–2729.
57. Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25(11):1451–1452.
58. Hudson RR (2000) A new statistic for detecting genetic differentiation. *Genetics* 155(4):2011–2014.
59. Liu C-W, et al. (2006) ATP binding and ATP hydrolysis play distinct roles in the function of 26S proteasome. *Mol cell* 24(1):39–50.
60. Rodriguez KA, Gaczynska M, Osmulski PA (2010) Molecular mechanisms of proteasome plasticity in aging. *Mech Ageing Dev* 131(2):144–155.

Supporting Information

Li et al. 10.1073/pnas.1514896112

SI Materials and Methods

Sample Collection. This study is a continuation of two earlier studies: (*i*) on possible sympatric speciation, the first report in subterranean mammals (4); and (*ii*) genome analysis of *S. galili* (8) ($2n = 52$). Two abutting but sharply contrasting soil types, white chalky rendzina and brownish volcanic basalt, are typical in two sites (Alma and Dalton) in central eastern Upper Galilee (31), Israel. There are four species of *Spalax* in Israel: *S. galili* ($2n = 52$), *S. golani* ($2n = 54$), *S. carmeli* ($2n = 58$), and *S. judaei* ($2n = 60$) (32). In this study, only *S. galili* was studied for whole genome analysis from the Alma Plio-Pleistocene basalt plateau vs. the Senonian chalk of Kerem Ben Zimra (11, 31). A total of 11 animals, 5 from chalk and 6 from basalt, were captured alive near Rehaniya, Upper Galilee (33.04E, 35.49N), in January 2014 (Fig. 1 *B* and *C*). After injecting Ketaset CIII at 5 mg/kg of body weight, animals were killed. Muscle tissues were isolated and stored in 95% (vol/vol) ethanol until further molecular analysis could be done.

Library Construction and Genome Sequencing. Genomic DNAs were isolated from the muscle samples of 11 individuals of *S. galili* using the DNeasy Blood and Tissue Kit (Qiagen) following the manufacturer's protocol. For each individual, 1 μg of genomic DNA was fragmented into 450–550 bp for libraries, with a 500-bp insert size, using the Covaris S220 system. Genome sequencing libraries were prepared with the TruSeq DNA Sample Preparation kit according to the manufacturer's protocol (Illumina). Briefly, the cleaved DNA fragments were end-repaired and A-tailed, pair-end adaptors were ligated to the fragments, adapter-ligated products of ∼500 bp were selected and purified on QIAquick spin columns (Qiagen), and purified products were enriched by 10 cycles of PCR amplification with Phusion DNA Polymerase. The purity and concentration of DNA were assessed and quantified using an Agilent 2100 Bioanalyzer (Agilent) and Nanodrop 2000 spectrophotometer (Nanodrop). Massive parallel sequencing was performed on the Illumina HiSeq 2000 platform, and 125-bp paired-end reads were generated.

Genome Mapping and SNP Calling. In total, 245.69 Gb of raw reads were generated. We removed low-quality reads, including those with >10 nucleotides aligned to the adapter sequences, those of putative PCR duplicates, those with average base quality <15, those with >50% having a base quality score <10, and those with >10% unidentified nucleotides (N). As a result, 241.76 Gb of clean reads were retained. These high-quality paired-end reads were mapped onto the reference genome of *S. galili* (GenBank accession no. AXCS00000000) using Burrows-Wheeler Aligner (BWA) v0.7.8 (33) with the "mem" option and default parameters. The best alignments were generated in the SAM format by SAMtools v1.1 (34) with the "rmdup" option. Finally, for each individual of the blind mole rats, ∼99.37% of reads were mapped to ∼97.91% (at least 1×) or ∼70.13% (at least 5×) of the reference genome assembly with 7.13-fold average depth.

After genome mapping, we undertook SNP calling for two abutting *S. galili* populations (chalk population consisting of five individuals and basalt population consisting of six individuals), using a Bayesian approach as implemented in SAMtools v1.1. The "mpileup" command was executed to identify SNPs with the parameters as "-q 20 -Q 20 -C 50 -t DP -m 2 -F 0.002." The probability of each probable genotype in a given SNP position was calculated with SAMtools, and the genotype with the highest posterior probability was picked. The low-quality SNPs were filtered by the Perl script vcfutils.pl in BCFtools v1.1 (34) package with the "varFilter" option and parameters as "-d 20 -D 140," and the high-quality SNPs [coverage depth ≥20 and ≤140, root mean square (RMS) of mapping quality ≥10, the distance of adjacent SNPs ≥5 bp] were retained for further analysis. Deviations from Hardy–Weinberg equilibrium (HWE) were tested by using Vcftools (vcftools.sourceforge.net/). SNPs deviating from HWE ($P < 0.05$) were excluded from subsequent analysis.

Verification by Sanger Sequencing. Twenty-two *OR*s, 20 *Tas2r*s, and 18 putatively neutral noncoding loci from 16 basalt and 13 chalk individuals of the blind mole rats were sequenced by traditional Sanger sequencing technology, aiming to identify genetic differentiation of chemosensory receptor genes between chalk and basalt populations (see details below). We also took advantage of these Sanger sequencing data (SSD) to estimate false-positive rates and false-negative rates in genome-wide SNP calling for our next-generation sequencing data (NSD). We divided SNP sites from SSD into three groups: group 1 contained SNPs that are absent between individuals of the blind mole rats in SSD but present in NSD; group 2 consisted of SNPs that are present between individuals in SSD but absent in NSD; and group 3 included SNPs that are present between individuals in both SSD and NSD. The three groups of SNPs were considered to be false positives (group 1), false negatives (group 2), and true positives (group 3). We identified 291 SNPs from SSD, whereas 273 SNPs were found in the equivalent loci from NSD. The three groups contained 16 (false positives), 34 (false negatives), and 257 (true positives) SNPs, respectively. Thus, our NSD was estimated to have a false-positive rate of 5.86% and a false-negative rate of 12.45%.

Population Structure Analysis. Population structures were investigated using three approaches. The first approach is the nonparametric PCA (35). PCA is widely used in population structure analysis because it is computationally efficient in handling large numbers of SNP markers (36, 37). The variant calling format was converted to binary ped format using VCFtools and PLINK v1.07 (pngu.mgh.harvard.edu/∼purcell/plink/) (38). PCA was performed using GCTA v1.24 (39) with the parameter "-pca 2." The second approach is to estimate individual ancestry based on the full maximum likelihood method as implemented in the program frappe v1.1 (40). Because the analysis is computationally intensive, we picked only the first SNP within each nonoverlapping window (window size: 20 kb) for the calculation. We assumed that the probable number of ancestral populations was 2, 3, and 4, respectively, according to the ecological information (4). The third approach is to infer population structures using a neighbor-joining method (41). The neighbor-joining phylogenetic tree was reconstructed using the nucleotide p-distance matrix, and the reliability of the tree was evaluated with 1,000 bootstraps in TreeBeST v1.9.2 (sourceforge.net/projects/treesoft/files/treebest/).

LD Analysis. To estimate the LD patterns between the two populations, we used the program Beagle v4.0 (42) to phase the genotypes into associated haplotypes with the command "gtgl." The correlation coefficient (r^2) between any two loci was calculated using VCFtools with the "hap-r2" option and default parameters. Average r^2 was calculated for pairwise SNPs in a 50-kb window with a custom written Perl script and was plotted against physical distance in base pairs with R v3.1.2 (43).

Genetic Diversity and Recombination Rate. Genetic diversity was estimated by Watterson's θ (44). For each population, we undertook a sliding window analysis, with a window size of 20 kb and a step

size of 5 kb. We calculated θ for each window, and the mean value of θ was considered as whole genome genetic diversity. The significance of difference in θ between two populations was tested with the Mann–Whitney U test.

The *interval* program in the LDhat v2.2 package (45) was used to infer the population recombination rate (ρ) across 28 scaffolds, each of which is longer than 10 Mb. For each population, the program was run with all of the SNPs from each scaffold. To reduce computational cost, a likelihood look-up table that assumes a population mutation rate (θ) of 0.001 was obtained from LDhat website (ldhat.sourceforge.net), and number of sequences was subsequently specified for each population with an *lkgen* program in the LDhat package. With a penalty parameter of 5 for changes in recombination rate, the modified look-up tables were applied to the *interval* program. Ten million iterations were run with the first 10% iterations discarded as burn-in.

Estimation of Demographic Parameters. We used the software Generalized Phylogenetic Coalescent Sampler (G-PhoCS) (46) to infer demographic parameters for the basalt and chalk populations including effective population size, migration rate, and population divergence time. G-PhoCS models gene flow with a Bayesian manner requiring the input of separate neutral loci and average mutation rate of neutral loci. A total of 3,000 putatively neutral loci with 1 kb in length were randomly picked along the genome, following three criteria: (*i*) these loci contain at least one SNP; (*ii*) they are at least 200 kb away from any known gene; and (*iii*) minimum interlocus distance of these loci is 50 kb. Migrations between two populations were assumed, whereas other parameters were set to the defaults. Sixty million iterations were run with the first 10% iterations discarded as burn-in. The convergence of Markov chain Monte Carlo (MCMC) algorithms was inspected by Tracer v1.6 (47). We repeated this analysis three times independently, all of which yielded similar results.

The demographic parameters generated by G-PhoCS are all scaled by the mutation rate. We estimated the average mutation rate $\mu = 3.03 \times 10^{-9}$ mutations per site per year, among rodents, by the average substitutions per site in neutral loci of 0.28 (48) divided by the average divergence time of 92.3 Mya (49). For each population, the effective population size (*Ne*) equals $\theta/4g\mu$, where g stands for generation time, and θ refers to population mutation rate. Based on a previous study (50), we assumed an average *S. galili* generation time of 1-y G-PhoCS estimated the mean population mutation rate of basalt population ($\theta_{\text{basalt}} = 1.418 \times 10^{-3}$), chalk population ($\theta_{\text{chalk}} = 8.784 \times 10^{-4}$), and ancestral population ($\theta_{\text{ancestor}} = 1.026 \times 10^{-3}$). Thus, the mean *Ne* of basalt was estimated to be 116,800, mean *Ne* of chalk was 72,380, and mean *Ne* of ancestral population was 84,570.

By specifying a source population (S) and a target population (T), G-PhoCS infers the migration rate from S to T as $m_{S\text{-}T}$, which can be converted into migrants from S to T per generation ($M_{S\text{-}T}$) with $M_{S-T} = m_{S-T} \times \theta_T$, where θ_T is the population mutation rate of the target population. We estimated the mean migration rate from basalt to chalk ($m_{b\text{-}c} = 6.613 \times 10^{3}$) and migration rate from chalk to basalt ($m_{c\text{-}b} = 2.036 \times 10^{3}$). Thus, the mean number of migrants per generation from basalt to chalk $M_{b\text{-}c}$ was 5.809, whereas the mean number of migrants per generation from chalk to basalt $M_{c\text{-b}}$ was 1.788.

The divergence time between chalk and basalt *S. galili* populations based on individual genome sequences was inferred using G-PhoCS. Using the average mutation rate $\mu = 3.03 \times 10^{-9}$, the average mutations per site (τ) between the chalk and basalt populations was estimated to be 6.917×10^{-4}. We estimated the divergence time (T) between the two populations by using the ratio of τ to μ. Thus, the mean of T was inferred to be 0.228 Mya, with a 95% confidence interval (CI) of [0.209, 0.250] Mya. For comparison, we also isolated the 13 mitochondrial protein-coding genes from the whole genomes of the 11 individuals of *S. galili* for dating. The equivalent genes from six outgroup species of mouse (*Mus musculus*; GenBank accession no.: KP090294), rat (*Rattus rattus*; GenBank accession no. KM577634), reed vole (*Microtus fortis*; GenBank accession no. JF261174), bamboo rat (*Rhizomys pruinosus*; GenBank accession no. KC789518), spiny mouse (*Acomys cahirinus*; GenBank accession no. JN571144), and a Middle East blind mole rat (*Spalax ehrenbergi*; GenBank accession no. AJ416891) were downloaded from National Center for Biotechnology Information (NCBI) (www.ncbi.nlm.nih.gov/). The 13 mitochondrial genes were concatenated and aligned. A Bayesian phylogenetic tree was constructed to evaluate the relationships of 11 individuals of *S. galili*. Divergence times were estimated using BEAST (51) with the Hasegawa–Kishino–Yano (HKY) model and an assumption of constant population size. By using the divergence time (27–34 Mya) (49) between rat and mouse as the calibration point, the divergence time of the basalt and chalk populations was estimated to have a mean of 0.425, and a 95% CI of [0.246, 0.645] Mya.

Putatively Selected Genes During Speciation. Tajima's D, calculated by dividing the difference between the mean number of pairwise nucleotide differences and the number of segregating sites by the square root of its SE, was used to test whether a sequence is under selection (52). The fixation index (F_{ST}) (53) was also chosen as an indicator of population differentiation. Tajima's D and F_{ST} values were calculated by Vcftools (vcftools.sourceforge.net/) with a window size of 20 kb and a step size of 5 kb.

Tajima's D and F_{ST} values were sorted in descending order separately. Windows that share the highest 5% of F_{ST} and lowest 5% Tajima's D estimates of each population were recognized as positively selected regions in a given population, and genes located in these regions are considered putatively selected genes.

Functional Enrichment Analysis. Functional enrichment analysis of GO terms was conducted using DAVID (Database for Annotation, Visualization, and Integrated Discovery) (54). The gene IDs of the reference genome of *S. galili* were converted to Uniprot IDs. Putatively selected genes were submitted to DAVID for enrichment analysis of the significant overrepresentation of GO terms, and the whole gene set of the *S. galili* genome was appointed as the background. We undertook Benjamini-corrected modified Fisher's exact tests to examine the significance of functional enrichment between various gene sets, and P values [i.e., Expression Analysis Systematic Explorer (EASE) scores] less than 0.05 were considered significant. GO biological processes (GO-BP), molecular function (GO-MF), and cellular component (GO-CC) terminologies were used.

Population Differentiation of Olfactory and Taste Receptor Genes. We identified olfactory receptor genes (*OR*s) and bitter taste receptor genes (*Tas2r*s) from the reference genome of *S. galili* using TblastN searches, with published mammalian *OR*s and *Tas2r*s as query sequences. Other taste receptor genes were not examined, because they evolve much slower than *Tas2r*s (55). We identified 1,042 *OR*s and 32 *Tas2r*s that are complete and intact and putatively functional. Two phylogenetic trees were reconstructed to show phylogenetic relationships of these genes (Figs. S2 and S3). To represent major lineages of the phylogenetic trees, we carefully chose 22 *OR*s and 20 *Tas2r*s for Sanger sequencing in 16 basalt and 13 chalk individuals of the blind mole rats. For comparison, we also sequenced 18 putatively neutral noncoding loci in the same 29 mole rats. These noncoding loci were chosen randomly from the genome, with a requirement that they are located at least 200 kb away from any known gene. These putatively neutral regions are all ~1,000 bp in length, which is similar to the average length of *OR*s or *Tas2r*s.

PCRs were conducted with high-fidelity KOD-Plus-Neo DNA polymerase (Toyobo) following the manufacturer's recommended procedures. PCR products were purified by the QIAquick PCR

Purification Kit (Qiagen) and subjected to direct sequencing in both strands with the same primers as used for PCRs. SNPs were identified by visually inspecting sequence chromatogram files. MEGA6 (56) was used to align and edit the sequences. Because mammalian *OR*s and *Tas2r*s are intron-less with ∼930 bp in length, we focused exclusively on coding regions for genetic analysis. Population genetic differentiation between the chalk and basalt populations was estimated as the fixation index F_{ST} using DnaSP v5.10 (57), and the significance level was determined by the Snn test (58). Pairwise F_{ST} statistics between basalt and chalk populations are provided in Table 1.

Whole Tissue Lysates, Proteasome Activity, and Western Blots. The muscle tissue from each animal (seven animals from each soil population) was weighed and disrupted in a 2-mL Potter–Elvehjem homogenizer in reticulocyte standard buffer (RSB) (10 mM Hepes, pH 6.2, 10 mM NaCl, and 1.4 mM $MgCl_2$) at a weight to volume of 1 g of tissue to 5 mL of buffer for muscle. An equal portion of the tissue was placed in a RSB buffer supplemented with the addition of 1 mM ATP, 0.5 mM DTT, and 5 mM $MgCl_2$ to help maintain intact 26S subassemblies (59) for peptidolytic assays. To the other half, 1 tablet/10 mL protease and phosphatase inhibitor minitablets were added (Thermo Fisher Scientific) for use in Western blots. To clear the lysate of debris, the homogenates were spun at 1,000 × *g* for 12 min, and the supernatant was then stored as 50-μL aliquots at −80 °C until further use. Protein content was measured using the Pierce BCA Protein Assay (Life Technologies).

In each assay, 20 μg of whole tissue or subfractionated lysates was incubated with 100 μM substrate specific for the type of proteasome activity. A saturating concentration of proteasome inhibitor *N*-(benzyloxycarbonyl) leucinyl-leucinyl-leucinal (MG132) was added to parallel samples. The difference of the fluorescence released with and without the inhibitor was used as a measure of the net peptidolytic activity of proteasome as previously described using model peptide substrates to represent cleavage after hydrophobic [chymotrypsin-like (ChTL)] residues, basic residues [trypsin-like (TL)], and acidic residues (PGPH) (60).

For Western blots, tissue lysates were separated using a 4–20% SDS/PAGE (Biorad Life Sciences) and transferred to nitrocellulose membranes (Biorad Life Sciences). The membranes were probed with antibodies against the following: α7 (mouse mAb, PW8110, 1:5K), RPT5 (mouse mAb, PW8310, 1:5K), beclin-1 (rabbit mAb, #3495P), LC3A/B (rabbit mAb, #4108P), ATG5 (rabbit mAb, #8540P), and ATG7 (rabbit mAb, #8558P) (Cell Signaling Technology). GAPDH (mouse mAb, G8795, 1:20K; Sigma-Aldrich) antibody (Thermo-Fisher Scientific) was used as a loading control. Primary antibodies were detected using anti-rabbit IRDye 680LT or anti-mouse IR Dye800 CW (Li-Cor) conjugated antibodies. Secondary antibodies were incubated at 1:10,000 for 2 h at room temperature, and images were captured and subsequently quantified using the Odyssey Imaging System (Li-Cor) by quantifying fluorescent signals as Integrated Intensities (I.I. K Counts) using the Odyssey Infrared Imaging System, Application Version 3.0 software. Prism 5 (GraphPad Software) was used to analyze and graph the various datasets generated in this study. Comparisons were analyzed using multiple *t* tests. Statistical significance was set at the $P < 0.05$ level with Holm–Sidak corrections to counteract the probability of false positives.

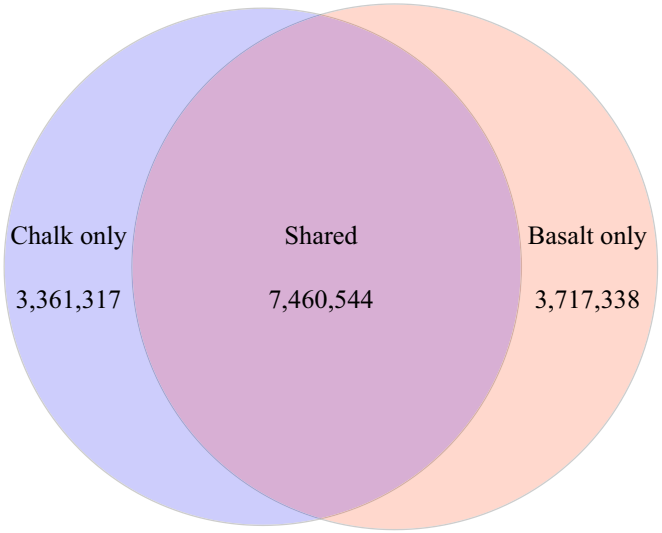

Fig. S1. Unique and shared SNPs between the chalk and basalt mole rat populations shown by Venn diagram.

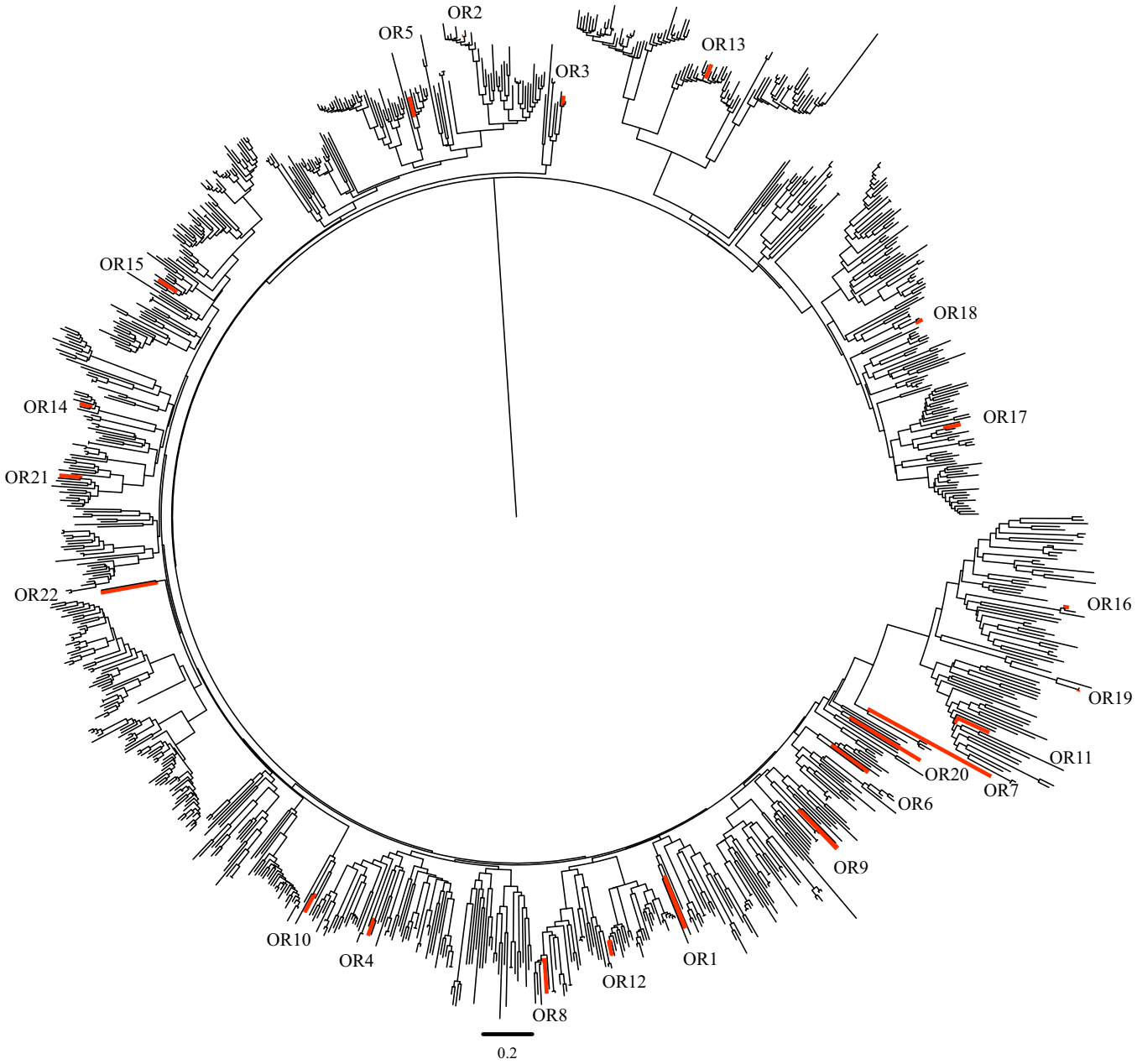

Fig. S2. Olfactory receptor genes analyses. Neighbor-joining phylogenetic tree of all 1,042 intact and complete olfactory receptor (OR) genes identified from the reference genome of *S. galili*. Red branches denote the 22 olfactory receptor genes surveyed in 29 blind mole rats from basalt and chalk populations. Nomenclature of the 22 OR genes was arbitrary.

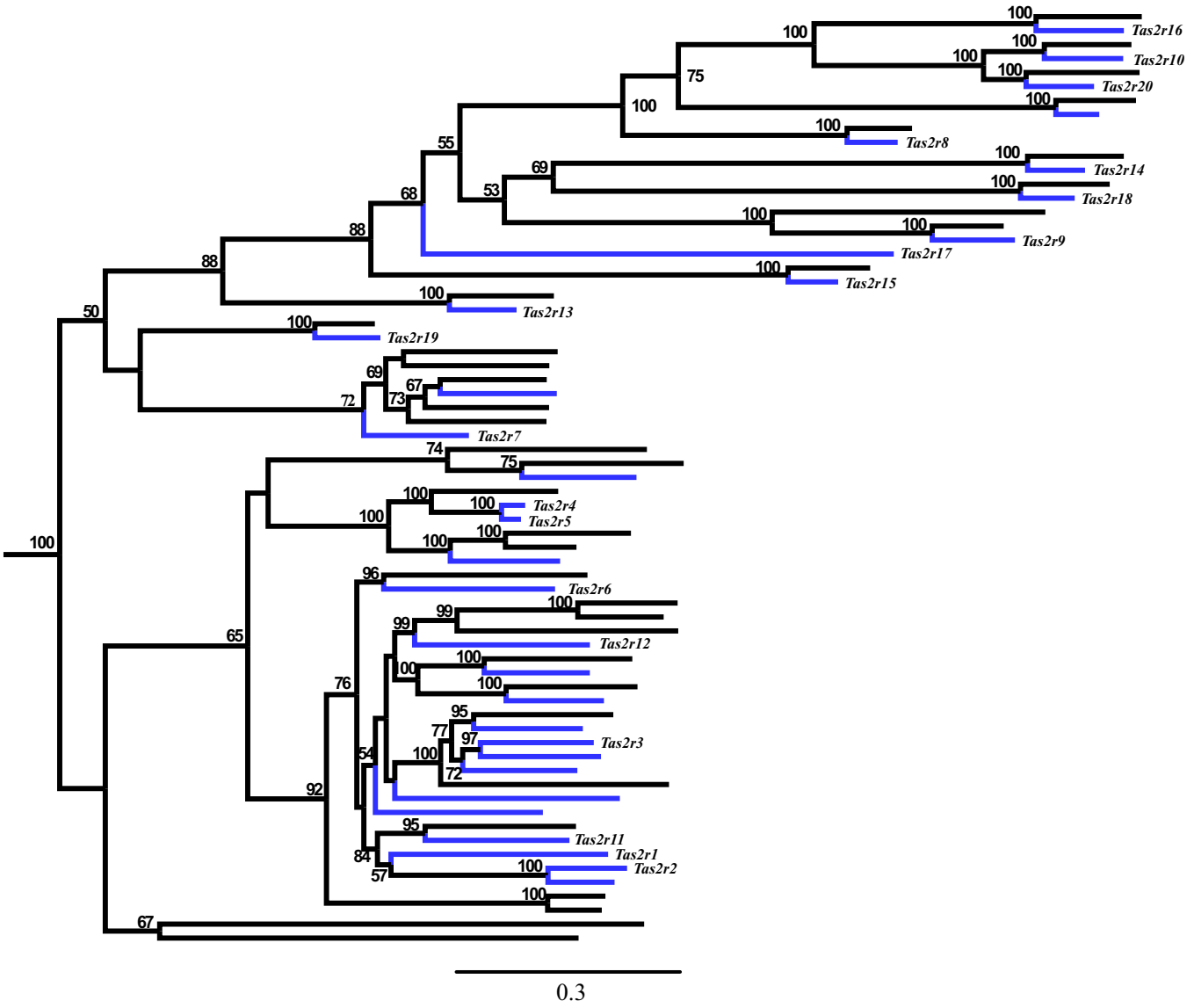

Fig. S3. Bayesian phylogenetic tree of 32 *S. galili* and 35 mouse *Tas2r* genes that are complete and intact. The numbers at the nodes are the Bayesian posterior probabilities shown as percentages. Blue branches indicate the 32 *Tas2r* genes from *S. galili*, whereas the remaining genes are from the mouse. Blue branches denote the 20 *Tas2r* genes surveyed in 29 individuals of *S. galili* from basalt and chalk populations. Nomenclature of the 20 *Tas2r* genes was arbitrary.

Other Supporting Information Files

SI Appendix (PDF)
Table S1 (DOCX)
Table S2 (DOCX)
Table S3 (DOCX)
Table S4 (DOCX)

Supplementary Note

Sympatric speciation, by definition, is the origin of new species within a freely interbreeding population. Recently, sympatric speciation has received widespread attention in the scientific community and has been the subject of heated debate[1-36].

1. Forbes AA, Powell TH, Stelinski LL, Smith JJ, & Feder JL (2009) Sequential sympatric speciation across trophic levels. *Science* 323(5915):776-779.
2. Rolán-Alvarez E (2007) Sympatric speciation as a by-product of ecological adaptation in the Galician Littorina saxatilis hybrid zone. *J Mollus Stud* 73(1):1-10.
3. Doebeli M & Dieckmann U (2003) Speciation along environmental gradients. *Nature* 421(6920):259-264.
4. Nosil P (2008) Speciation with gene flow could be common. *Mol Ecol* 17(9):2103-2106.
5. Papadopulos AS*, et al.* (2011) Speciation with gene flow on Lord Howe Island. *Proc Natl Acad Sci USA* 108(32):13188-13193.
6. Martin CH (2013) Strong assortative mating by diet, color, size, and morphology but limited progress toward sympatric speciation in a classic example: Cameroon crater lake cichlids. *Evolution* 67(7):2114-2123.
7. Jones AG, Moore GI, Kvarnemo C, Walker D, & Avise JC (2003) Sympatric speciation as a consequence of male pregnancy in seahorses. *Proc Natl Acad Sci USA* 100(11):6598-6603.
8. Friesen V*, et al.* (2007) Sympatric speciation by allochrony in a seabird. *Proc Natl Acad Sci USA* 104(47):18589-18594.
9. Gavrilets S & Waxman D (2002) Sympatric speciation by sexual conflict. *Proc Natl Acad Sci USA* 99(16):10533-10538.
10. Higashi M, Takimoto G, & Yamamura N (1999) Sympatric speciation by sexual selection. *Nature* 402(6761):523-526.
11. Van Leuven JT, Meister RC, Simon C, & McCutcheon JP (2014) Sympatric speciation in a bacterial endosymbiont results in two genomes with the functionality of one. *Cell* 158(6):1270-1280.
12. Crow KD, Munehara H, & Bernardi G (2010) Sympatric speciation in a genus of marine reef fishes. *Mol Ecol* 19(10):2089-2105.
13. Via S (2001) Sympatric speciation in animals: the ugly duckling grows up. *Trends Ecol Evol* 16(7):381-390.
14. Barluenga M, Stölting KN, Salzburger W, Muschick M, & Meyer A (2006) Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature* 439(7077):719-723.
15. Savolainen V*, et al.* (2006) Sympatric speciation in palms on an oceanic island. *Nature* 441(7090):210-213.
16. Berlocher SH & Feder JL (2002) Sympatric speciation in phytophagous insects: moving beyond controversy? *Annu Rev Entomol* 47(1):773-815.
17. Bird CE, Fernandez-Silva I, Skillings DJ, & Toonen RJ (2012) Sympatric speciation in the post "modern synthesis" era of evolutionary biology. *Evol Biol* 39(2):158-180.
18. Bolnick DI (2011) Sympatric speciation in threespine stickleback: why not? *Int J Ecol* 2011.
19. Bolnick DI & Fitzpatrick BM (2007) Sympatric speciation: models and empirical evidence. *Annu Rev Ecol Evol*:459-487.
20. Schliewen UK, Tautz D, & Pääbo S (1994) Sympatric speciation suggested by monophyly of crater lake cichlids. *Nature* 368(6472):629-632.
21. Friedman J, Alm EJ, & Shapiro BJ (2013) Sympatric speciation: when is it possible in bacteria? *PLoS ONE* 8(1):e53539.
22. Jiggins CD (2006) Sympatric speciation: why the controversy? *Curr Biol* 16(9):R333-R334.
23. Feder JL, Egan SP, & Nosil P (2012) The genomics of speciation-with-gene-flow. *Trends Genet* 28(7):342-350.
24. Martin CH (2012) Weak disruptive selection and incomplete phenotypic divergence in two classic examples of sympatric speciation: Cameroon Crater Lake cichlids. *Am Nat* 180(4):E90-E109.
25. Fitzpatrick BM, Fordyce J, & Gavrilets S (2008) What, if anything, is sympatric speciation? *J Evol Biol* 21(6):1452-1459.
26. Michel AP*, et al.* (2010) Widespread genomic divergence during sympatric speciation. *Proc Natl Acad Sci USA* 107(21):9724-9729.
27. Soria-Carrasco V*, et al.* (2014) Stick insect genomes reveal natural selection's role in parallel

speciation. *Science* 344(6185):738-742.
28. Débarre F (2012) Refining the conditions for sympatric ecological speciation. *J Evol Biol* 25(12):2651-2660.
29. Johannesson K (2001) Parallel speciation: a key to sympatric divergence. *Trends Ecol Evol* 16(3):148-153.
30. Dieckmann U & Doebeli M (1999) On the origin of species by sympatric speciation. *Nature* 400(6742):354-357.
31. Kondrashov AS & Kondrashov FA (1999) Interactions among quantitative traits in the course of sympatric speciation. *Nature* 400(6742):351-354.
32. Wilson AB, Noack–Kunnmann K, & Meyer A (2000) Incipient speciation in sympatric Nicaraguan crater lake cichlid fishes: sexual selection versus ecological diversification. *Phil Trans R Soc B* 267(1458):2133-2141.
33. Gourbiere S (2004) How do natural and sexual selection contribute to sympatric speciation? *J Evol Biol* 17(6):1297-1309.
34. Papadopulos AS*, et al.* (2014) Evaluation of genetic isolation within an island flora reveals unusually widespread local adaptation and supports sympatric speciation. *Phil Trans R Soc B* 369(1648):20130342.
35. Boomsma JJ & Nash DR (2014) Evolution: sympatric speciation the eusocial way. *Curr Biol* 24(17):R798-R800.
36. Les DH*, et al.* (2015) Through thick and thin: Cryptic sympatric speciation in the submersed genus Najas (*Hydrocharitaceae*). *Mol Phylogenet Evol* 82:15-30.