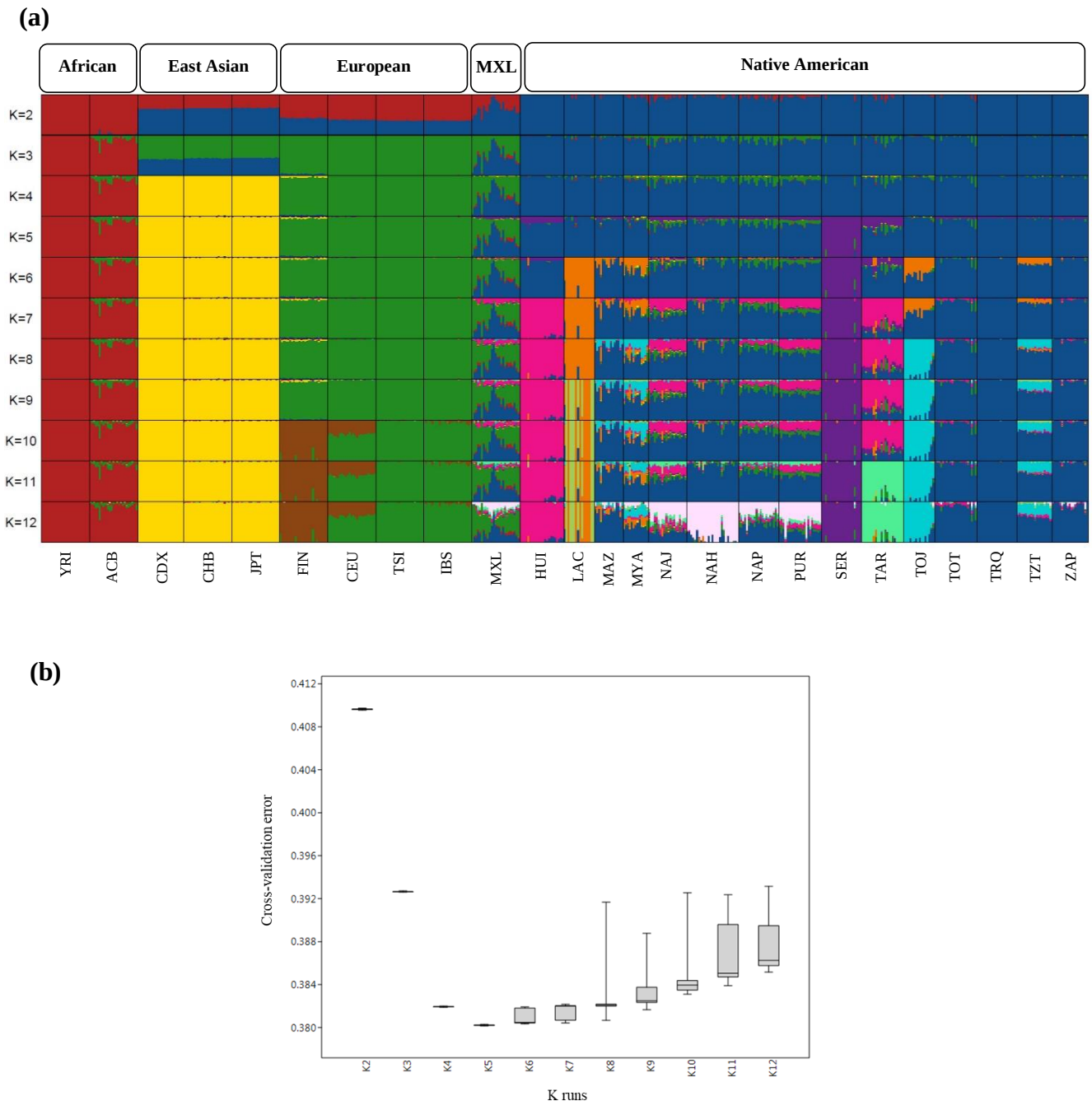
Dietary, Cultural and Pathogens-related Selective Pressures Shaped Differential Adaptive Evolution Among Native Mexican Populations

Claudia Ojeda-Granados, Paolo Abondio, Alice Setti, Stefania Sarno, Guido Alberto Gnecchi-Ruscone, Eduardo González-Orozco, Sara De Fanti, Andres Jiménez-Kaufmann, Héctor Rangel-Villalobos, Andrés Moreno-Estrada, Marco Sazzini*

*Author for correspondence: Marco Sazzini, Dept. of Biological, Geological and Environmental Sciences, University of Bologna, Italy, +39 051 2094314, e-mail: marco.sazzini2@unibo.it

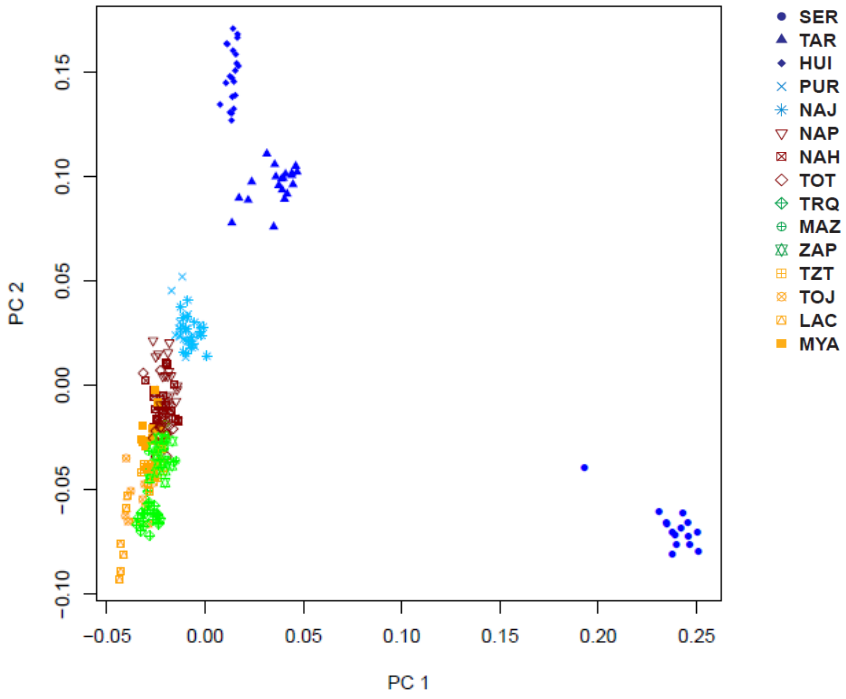
Supplementary Material

Includes supplementary tables S1-S6 (supplementary table S7 and S8 are reported in a separate Excel file), supplementary figures S1-S13, supplementary results and supplementary references.

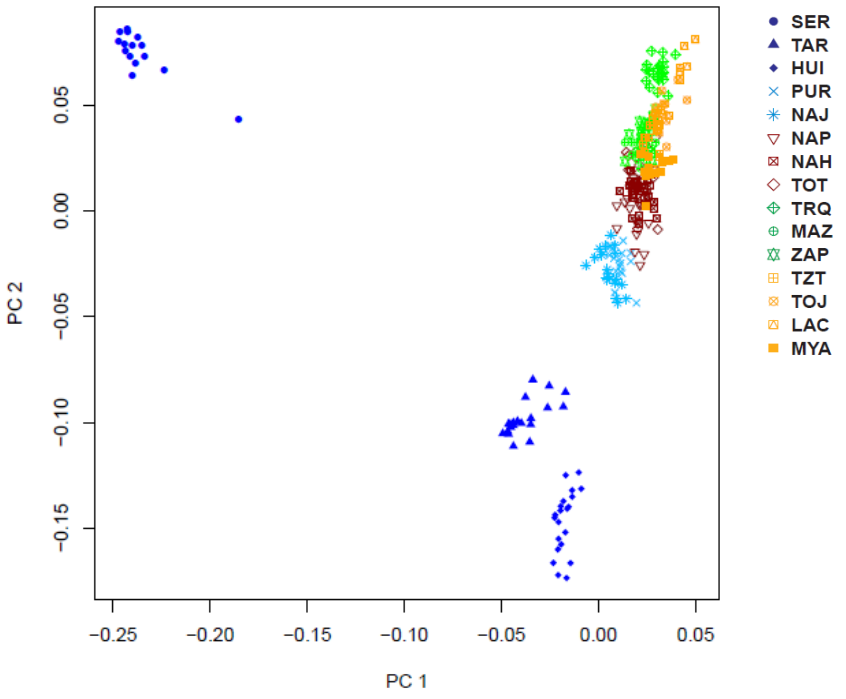


Supplementary figure S1. (a) Results of ADMIXTURE unsupervised clustering analyses on the 15 Native Mexican groups included in the NMDP and on worldwide populations of African, East Asian, European, and Mexican-American (MXL) ancestry. Ancestry proportions were estimated from K = 2 (top) to K = 12 (bottom). Individuals are grouped at population level. (b) Cross-validation (CV) errors computed for all the 25 independent runs performed for each of the tested K. The best predictive accuracy (i.e. lowest CV error) was achieved by the model when five ancestral components (K = 5) were tested.

(a)

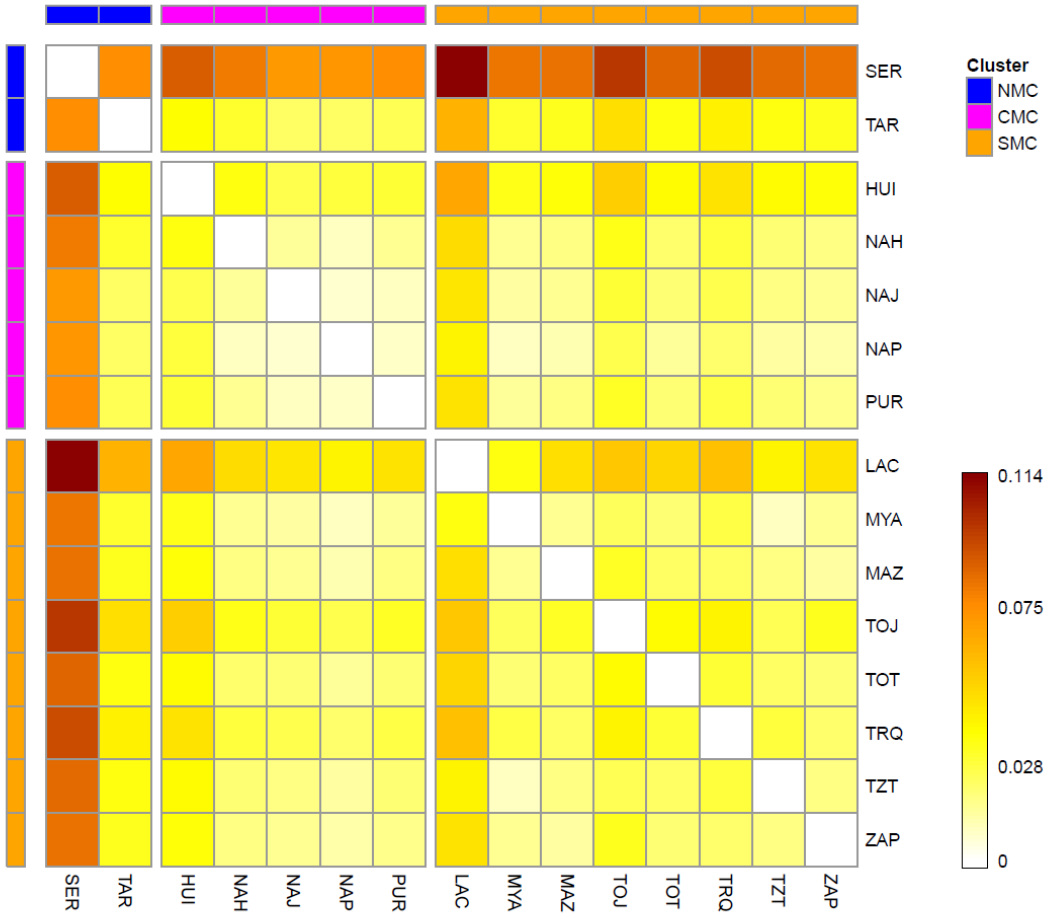


(b)

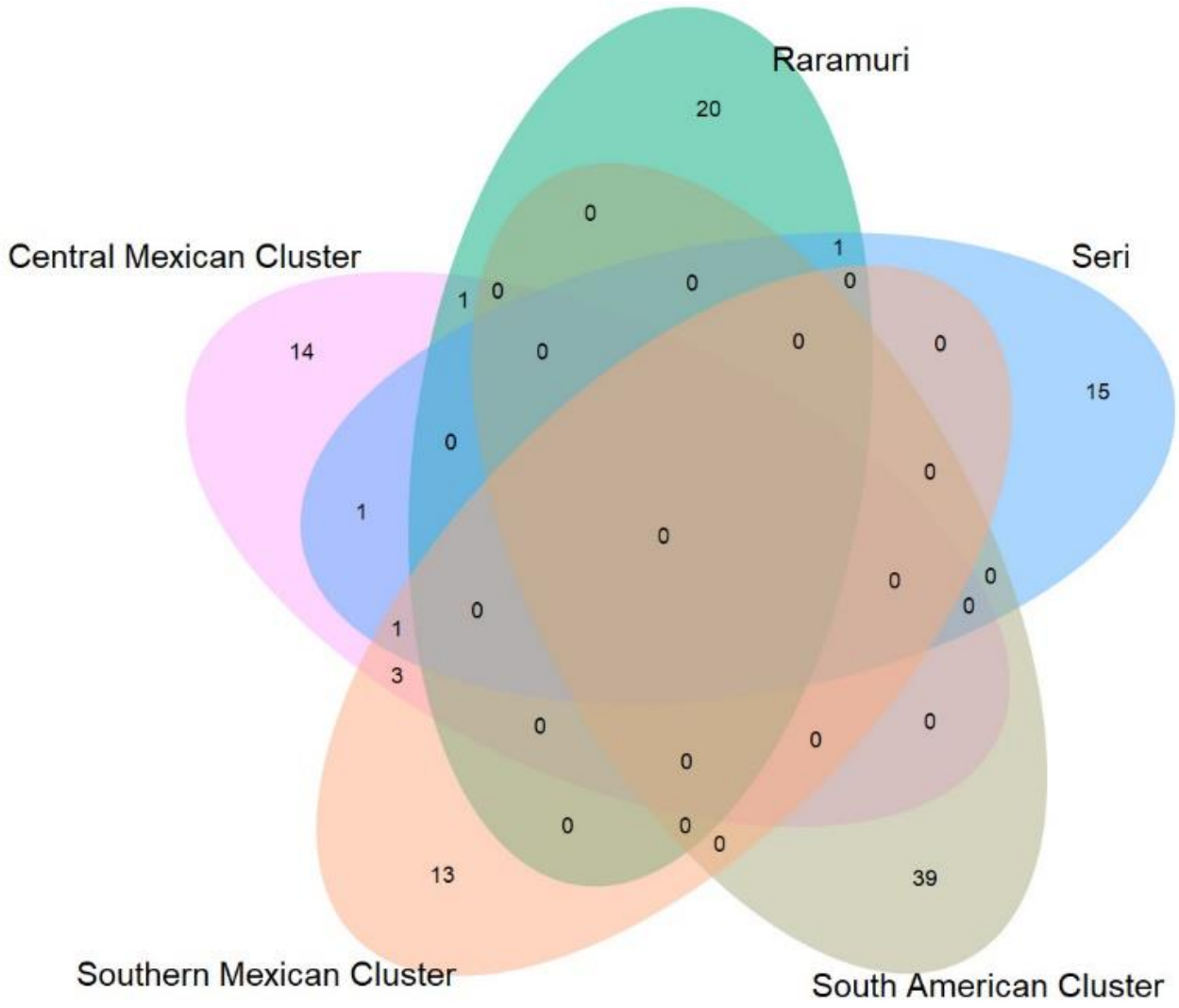


Supplementary figure S2. Comparison between Principal Component Analysis (PCA) performed on **a)** the non-imputed NMDP dataset and **b)** the imputed dataset including 249 individuals belonging to the 15 considered Native Mexican groups. In order to better compare patterns of population structure before

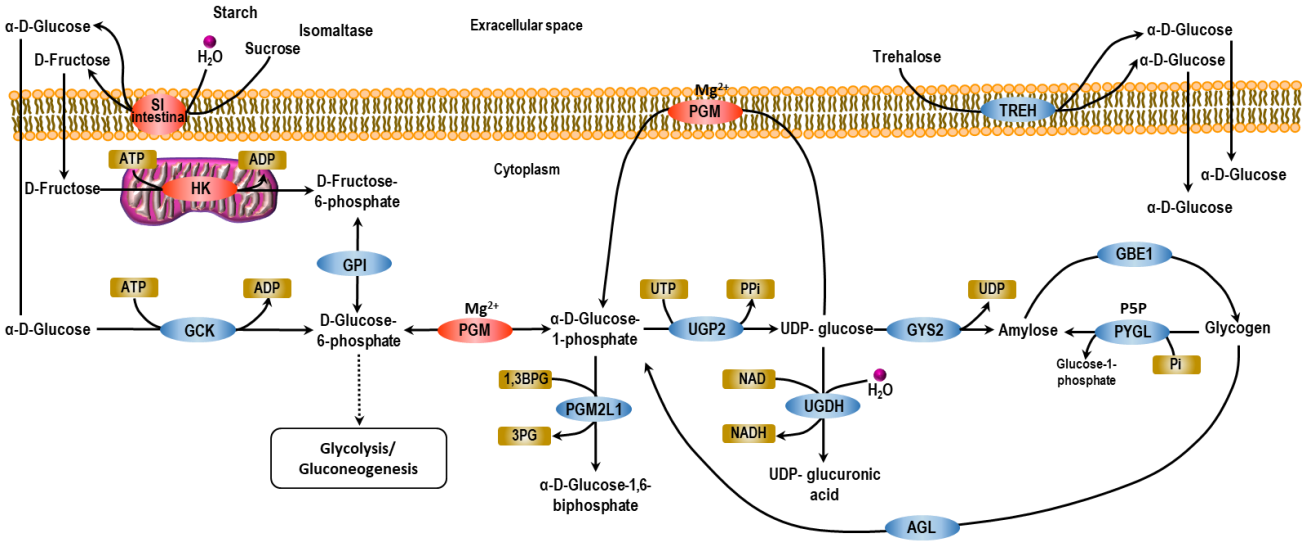
and after the imputation procedure, populations are color-coded according to their geographical locations as reported in Moreno-Estrada et al. 2014: North Mexico (dark blue), Central West Mexico (blue), Central East Mexico (red), South Mexico (green), and Southeast Mexico (yellow). High and significant correlation between genomic information summarized by the two most informative PCs computed on the non-imputed and imputed datasets was observed (correlation in a symmetric Procrustes rotation = 0.997, $p < 10^{-4}$). The distribution of genetic variation observed for the imputed NMDP dataset indeed resembles that depicted for the non-imputed one according to both PCA and ADMIXTURE analyses (supplementary figure S1, Supplementary Material online). In particular, SER diverged from the bulk of Mexican indigenous populations due to their prolonged genetic isolation and along with LAC represented respectively the northernmost and southernmost groups located at the edges of the observed northwest-southeast cline of variation. HUI were found to present some genetic affinity with northern Mexican populations, as supported also by ADMIXTURE analyses pointing to their appreciable proportion of the SER-like ancestry fraction (supplementary figure S1, Supplementary Material online) and by their branching as a distinct sub-group within the Central Mexican cluster according to fine STRUCTURE analysis (fig. 2).



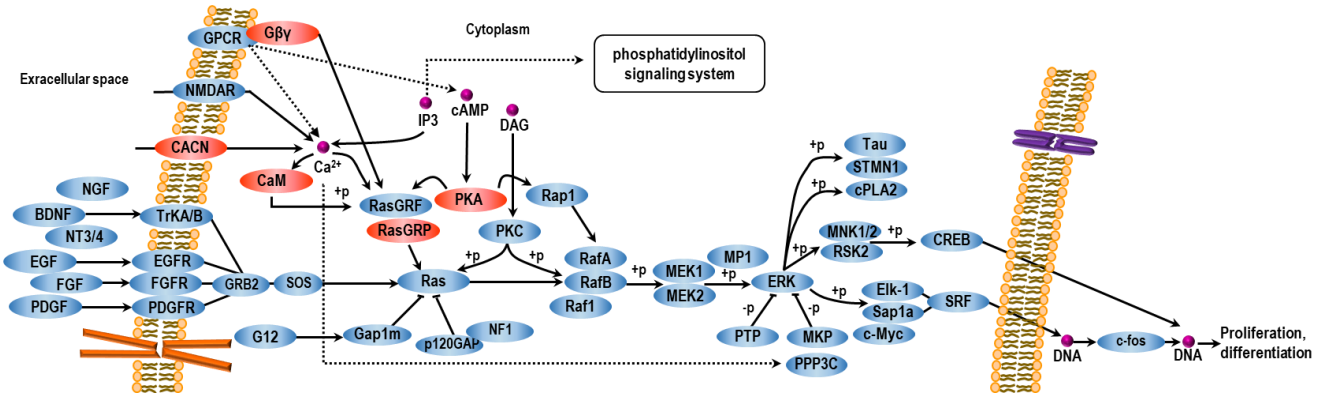
Supplementary figure S3. Heatmap of pairwise F_{st} values computed for the 15 considered Native Mexican populations. A color scale ranging from white to yellow and orange has been used to approximate lower-to-higher genetic distances, as summarized in the legend at the bottom right of the plot. To facilitate comparisons among populations both between and within different clusters, they have been grouped according to results from fineSTRUCTURE analysis, as detailed by the top and left colored bars of the plot, as well as by the legend at the upper right of it: Northern Mexican cluster, NMC (blue), Central Mexican cluster, CMC (magenta), Southern Mexican cluster, SMC (orange).



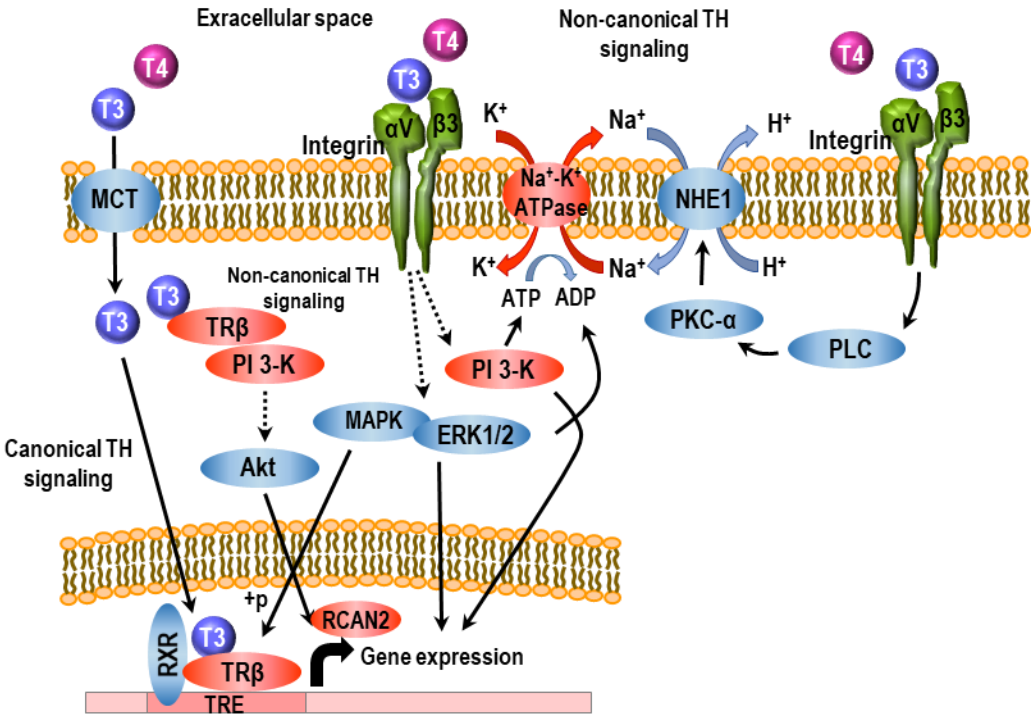
Supplementary figure S4. Venn diagrams illustrating the number of putative adaptive genes belonging to the gene networks detailed in supplementary tables S2-S6 that are shared among the examined Native Mexican population clusters, as well as between them and the South American control cluster. The Central Mexican cluster (CMC) was found to share three putative adaptive genes (*PLCB1*, *GNAI2* and *RHOA*) with the Southern Mexican cluster (SMC), and one gene (*PRKACB*) with both Seri and SMC. Rarámuri (TAR) and CMC shared one putative adaptive gene (*STAT1*). In addition, Seri shared one putative adaptive gene (*ATP1A2*) with Rarámuri and one (*GNG2*) with CMC. No putative adaptive genes were found to be shared between Mexican and South American clusters. Venn diagrams were plotted using the VennPainter tool (Lin et al. 2016).



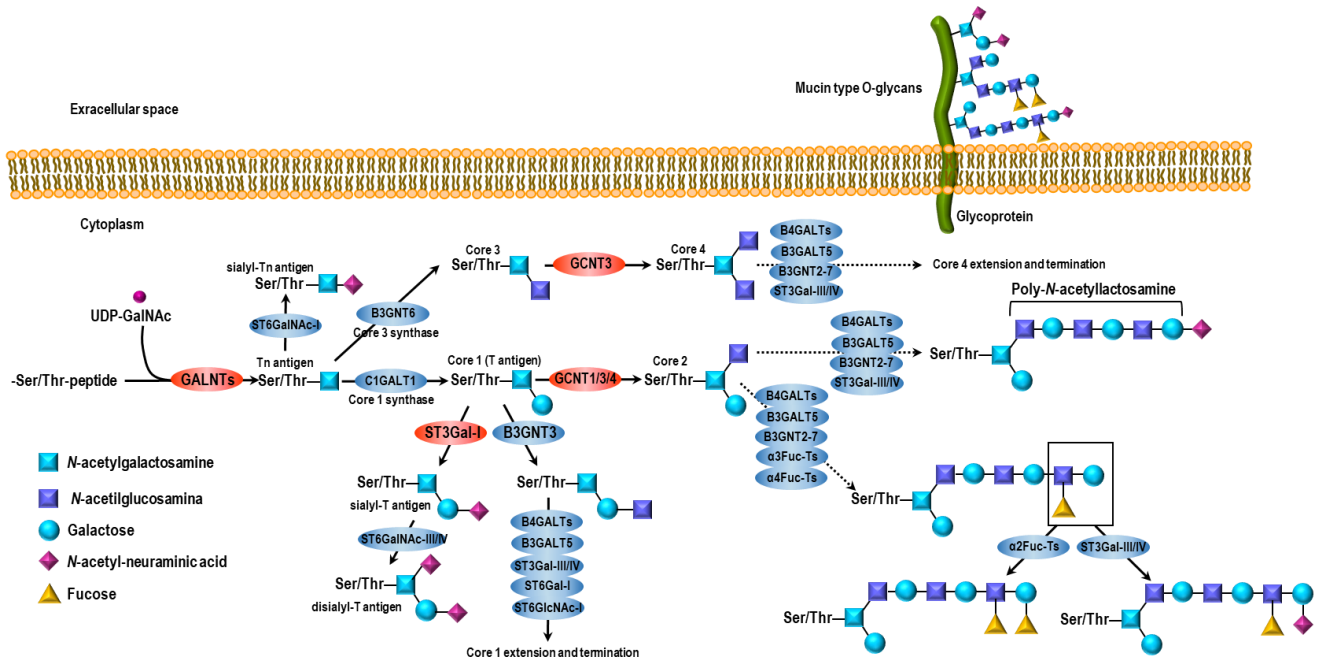
Supplementary figure S5. Representation of the *Starch and sucrose metabolism* pathway supposed to have played a role in biological adaptations evolved by the Seri population. Proteins encoded by the genes putatively targeted by natural selection and constituting one of the gene networks described in fig. 3a are displayed in red. SI is encoded by the *SI* gene; HK by *HK2*; PGM by *PGM1*.



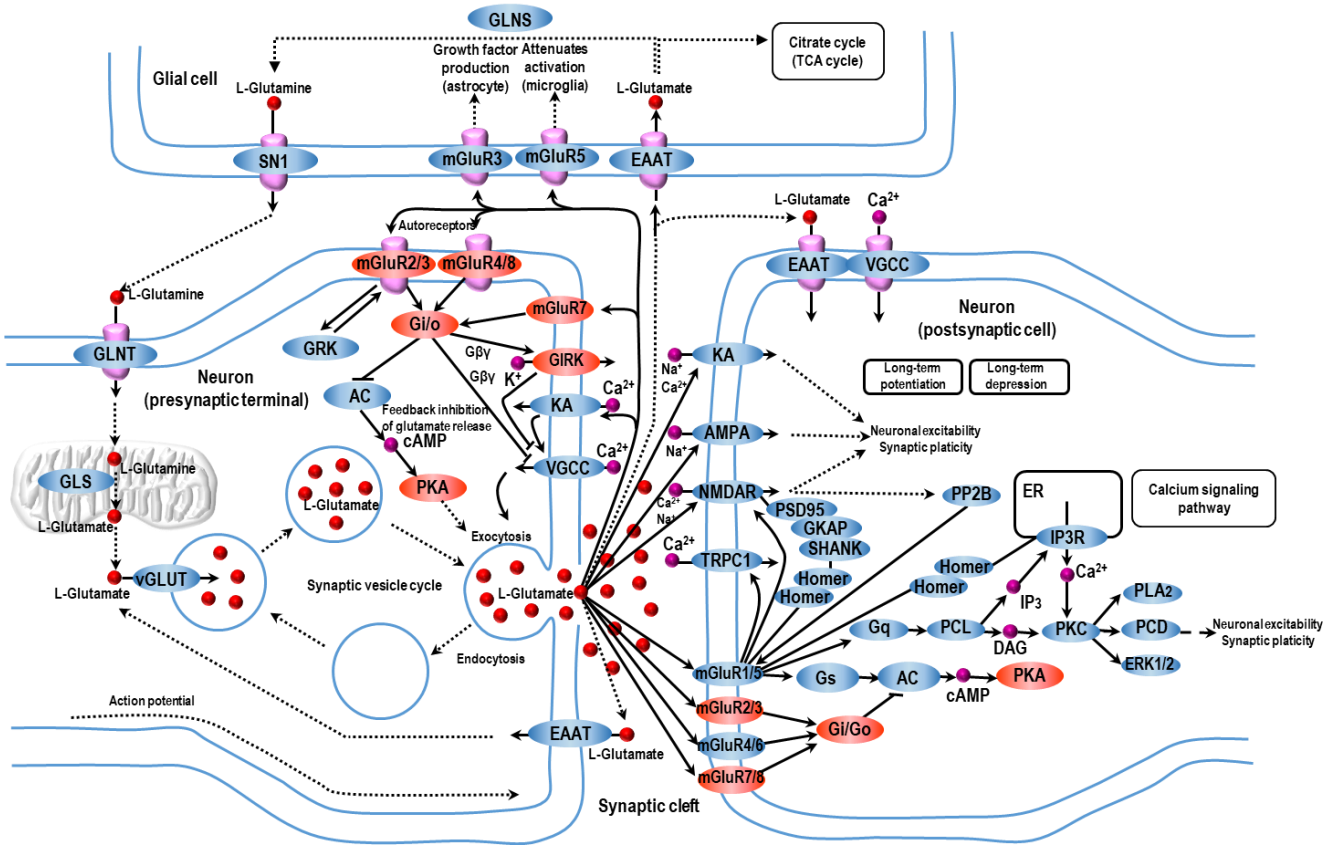
Supplementary figure S6. Integrated representation of the *Ras signaling* and canonical *MAPK signaling* pathways supposed to have played a role in biological adaptations evolved by the Seri population. Proteins encoded by the genes putatively targeted by natural selection and constituting two of the gene networks described in fig. 3a are displayed in red. CACN is encoded by the *CACNA1E*, *CACNA1D*, and *CACNA2D1* genes; RasGRP by *RASGRP3*; Gβγ by *GNG2;* CaM by *CALM1*; PKA by *PRKACB*.



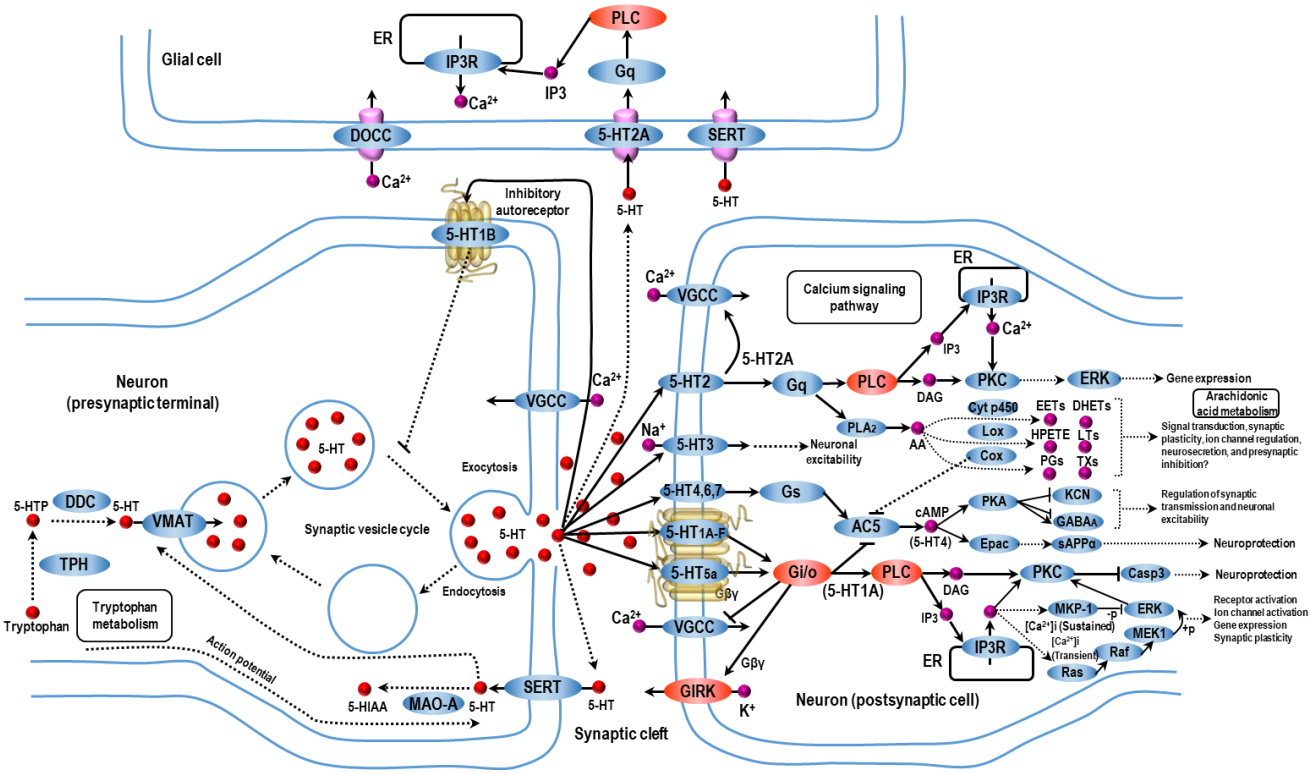
Supplementary figure S7. Representation of the *Thyroid hormone signaling* pathway supposed to have played a role in biological adaptations evolved by the Rarámuri population. Proteins encoded by the genes putatively targeted by natural selection and constituting one of the gene networks described in fig. 3b are displayed in red. Na^+/K^+-ATPase is encoded by the *ATP1A2* and *ATP1B1* genes; TRβ by *THRB*; PI 3-K by *PIK3R5*; RCAN2 by *RCAN2*.



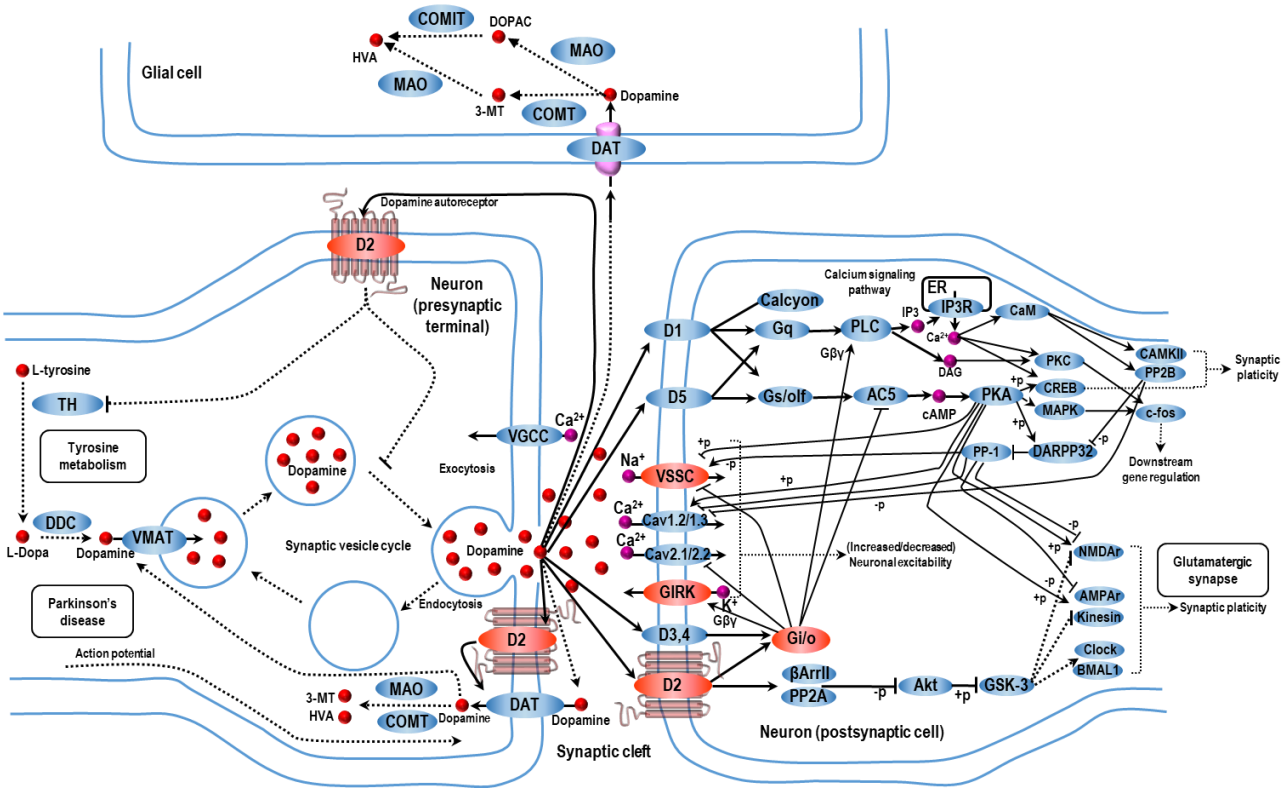
Supplementary figure S8. Representation of the *Mucin type O-Glycan biosynthesis* pathway supposed to have played a role in biological adaptations evolved by the Rarámuri population. Proteins encoded by the genes putatively targeted by natural selection and constituting one of the gene networks described in fig. 3b are displayed in red. GALNTs are encoded by the *GALNT2*, *GALNT6*, *GALNT7*, *GALNT10*, *GALNT14*, *GALNT15*, *GALNT16*, *GALNTL5*, *GALNTL6*, and *GALNT18* genes; GCNT3 by *GCNT3*; GCNT1/3/4 by *GCNT1* and *GCNT3*; ST3Gal-I by *ST3GAL1*.



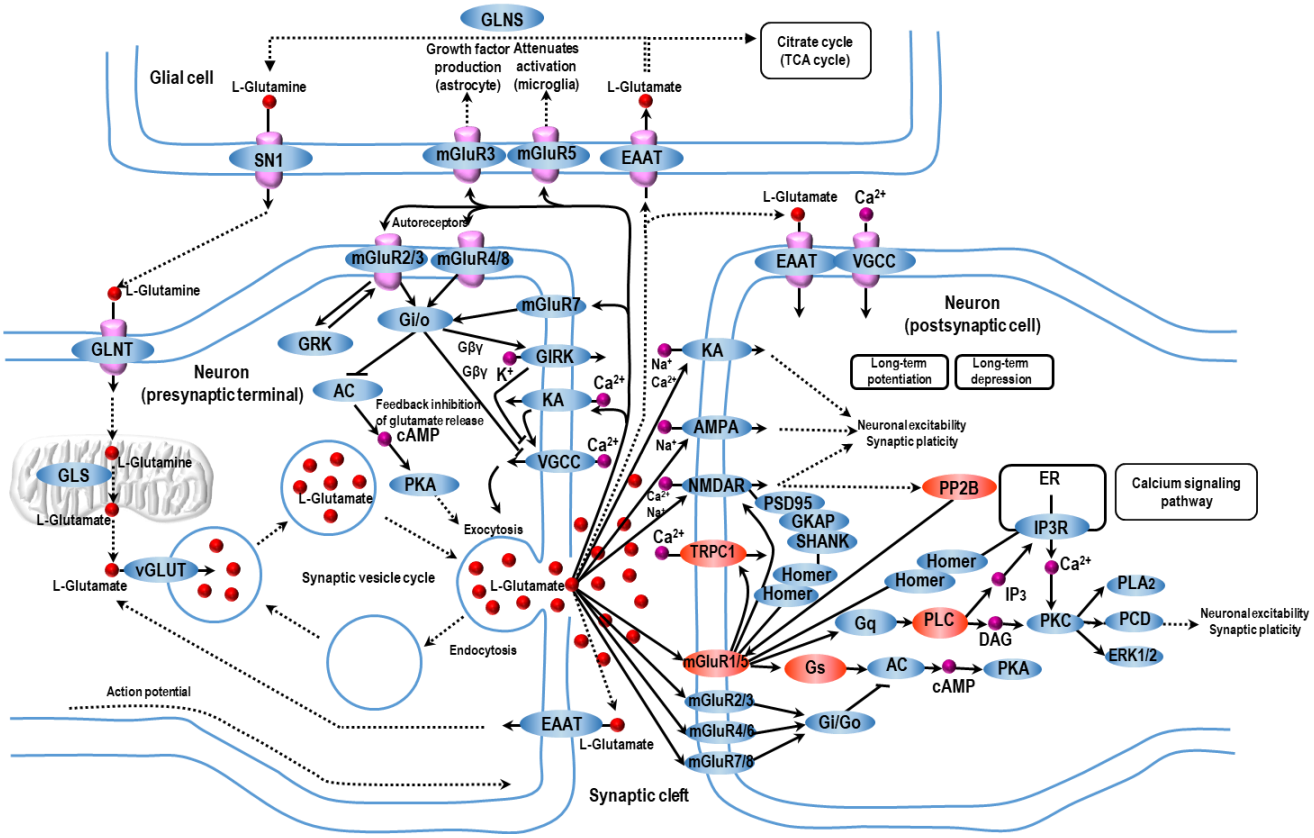
Supplementary figure S9. Representation of the *Glutamatergic synapse* pathway supposed to have played a role in biological adaptations evolved by populations belonging to the Central Mexican Cluster. Proteins encoded by the genes putatively targeted by natural selection and constituting one of the gene networks described in fig. 3c are displayed in red. mGluR2/3 is encoded by the *GRM2* gene; mGluR4/8 by *GRM8*; mGluR7 by *GRM7*; mGluR7/8 by *GRM7*, *GRM8*; Gi/o by *GNAI2*, *GNG2*, *GNG4*, *GNG5*, *GNG7*, *GNG10*; GIRK by *KCNJ3*; PKA by *PRKACB*.



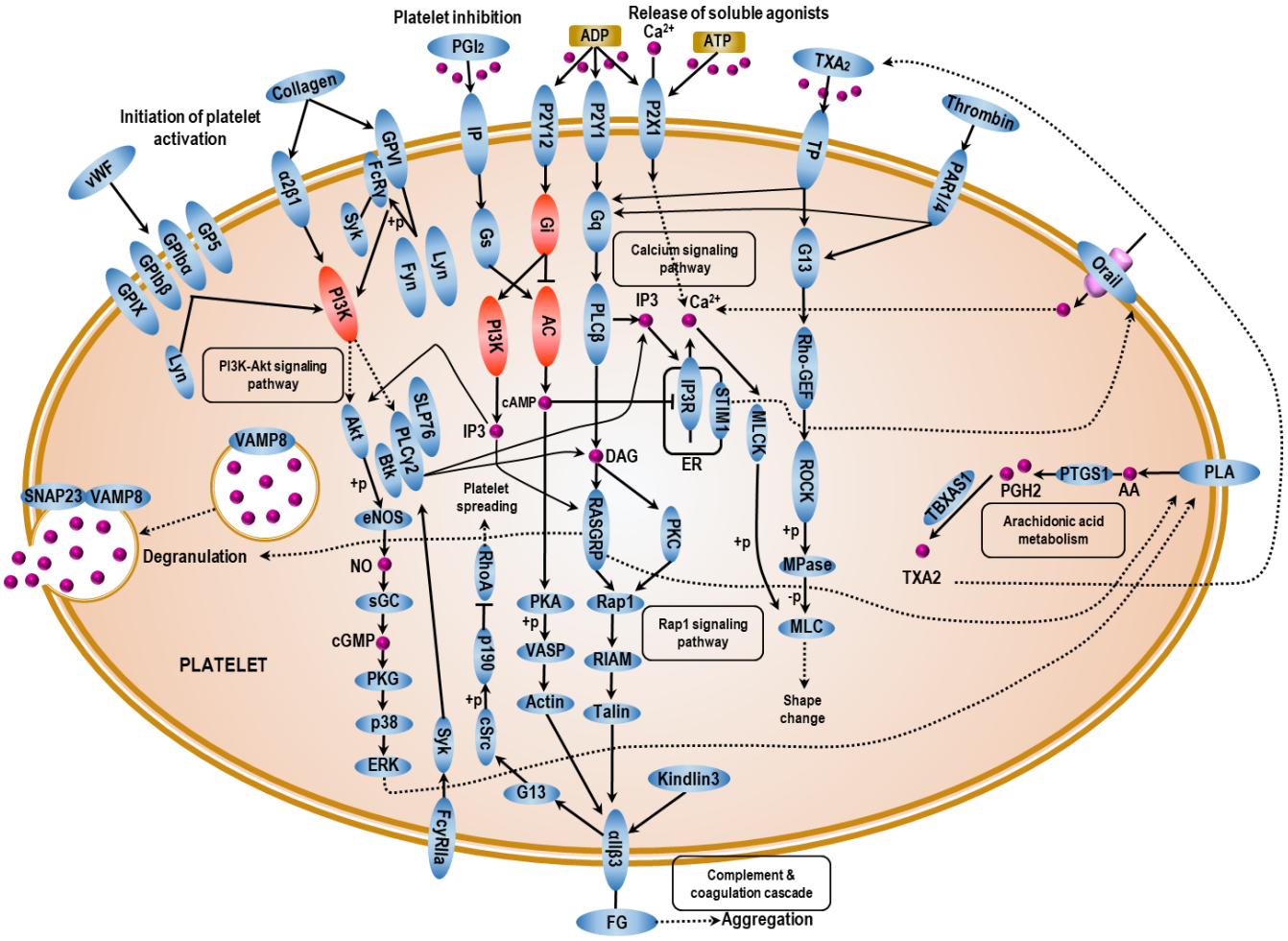
Supplementary figure S10. Representation of the *Serotonergic synapse* pathway supposed to have played a role in biological adaptations evolved by populations belonging to the Central Mexican Cluster. Proteins encoded by the genes putatively targeted by natural selection and constituting one of the gene networks described in fig. 3c are displayed in red. PLC is encoded by *PLCB1* and *PLCB4*; GIRK by *KCNJ3*; Gi/o by *GNAI2*, *GNG2*, *GNG4*, *GNG10*, *GNG12*, and *GNB5*.



Supplementary figure S11. Representation of the *Dopaminergic synapse* pathway supposed to have played a role in biological adaptations evolved by populations belonging to the Central Mexican Cluster. Proteins encoded by the genes putatively targeted by natural selection and constituting one of the gene networks described in fig. 3c are displayed in red. D2 is encoded by *DRD2*; VSSC by *SCN1A*; GIRK by *KCNJ3*; Gi/o by *GNAI2*, *GNG2*, *GNG4*, *GNG7*, and *GNB5*.



Supplementary figure S12. Representation of the *Glutamatergic synapse* pathway supposed to have played a role in biological adaptations evolved by populations belonging to the Southern Mexican Cluster. Proteins encoded by the genes putatively targeted by natural selection and constituting one of the gene networks described in fig. 3d are displayed in red. TRPC1 is encoded by *TRPC1*; mGluR1/5 by *GRM1*; Gs by *GNAS*; PLC by *PLCB1*; PP2B by *PPP3CA*.



Supplementary figure S13. Representation of the *Platelet activation* pathway supposed to have played a role in biological adaptations evolved by populations belonging to the Southern Mexican Cluster. Proteins encoded by the genes putatively targeted by natural selection and constituting one of the gene networks described in fig. 3d are displayed in red. PI3K is encoded by *PIK3CB*; Gi by *GNAI2*; AC by *ADCY3* and *ADCY8*.

Supplementary table S1. Detailed information for the 15 Native Mexican groups included in the "high-density imputed dataset" made up of 271 individuals.

Population	ID	N	Latitude	Longitude	Reference
Seri	SER	16	29.00	-112.15	NMDP (Moreno-Estrada et al. 2014)
Rarámuri	TAR	21	27.75	-107.17	NMDP (Moreno-Estrada et al. 2014)
Huichol	HUI	21	21.17	-104.08	NMDP (Moreno-Estrada et al. 2014)
Nahua (Jalisco)	NAJ	19	19.50	-103.50	NMDP (Moreno-Estrada et al. 2014)
Purepecha	PUR	21	19.75	-101.50	NMDP (Moreno-Estrada et al. 2014)
Totonac	TOT	19	20.00	-97.80	NMDP (Moreno-Estrada et al. 2014)
Nahua (Puebla)	NAP	20	19.97	-97.62	NMDP (Moreno-Estrada et al. 2014)
Nahua (Trios)	NAH	26	19.93	-97.62	NMDP (Moreno-Estrada et al. 2014)
Triqui	TRQ	21	17.18	-97.95	NMDP (Moreno-Estrada et al. 2014)
Zapotec (South)	ZAP	19	17.23	-96.23	NMDP (Moreno-Estrada et al. 2014)
Mazatec	MAZ	19	18.33	-96.33	NMDP (Moreno-Estrada et al. 2014)
Tzotzil	TZT	16	16.83	-92.67	NMDP (Moreno-Estrada et al. 2014)
Tojolabal	TOJ	11	16.50	-92.00	NMDP (Moreno-Estrada et al. 2014)
Lacandon	LAC	11	16.75	-91.25	NMDP (Moreno-Estrada et al. 2014)
Maya (Q.R.)	MYA	11	19.58	-88.58	NMDP (Moreno-Estrada et al. 2014)

ID, population label; N, number of samples after the applied pre- and post- imputation quality control filtering; NMDP, Native Mexican Diversity Panel; Q.R., Quintana Roo.

Supplementary table S2. Gene networks showing significant signatures of positive selection in the control South American cluster according to *signet* analyses performed on the obtained genome-wide distributions of nSL and H12 scores.

Pathway	P. size	N. size	HSS	*p*-value	Genes
PI3K-Akt signaling (nSL)	305	14	6.574	0.002	***CSF1R EFNA5*** *ERBB4* ***FGF1 FGF2 FGFR2 FLT1*** *FLT3* ***ANGPT1 IGF1R*** *IRS1 NTRK2 BDNF* ***FGF18***
Rap1 signaling (nSL)	190	14	6.500	0.003	***CSF1R EFNA5 FGF1 FGF2 FGFR2 FLT1*** *FGF20* ***ANGPT1*** *IGF1* ***IGF1R*** *PDGFRB PDGFC TEK* ***FGF18***
Human papillomavirus infection (nSL)	328	12	6.353	0.003	*ITGA11 FN1 COL6A3 TNC IBSP LAMA1 LAMA4 LAMC1 LAMC2 SPP1 THBS4 TNR*
Arachidonic acid metabolism (H12)	45	3	5.246	0.001	*CYP2C8* ***GPX5 GPX6***
Glutathione metabolism (H12)	57	3	8.441	0.003	*GGT5* ***GPX5 GPX6***
Staphylococcus aureus infection (H12)	74	4	2.001	0.004	*C1QB C2 C3 CFB*

P. size, number of genes belonging to the pathway; N. size, number of genes composing the identified network; HSS, highest scoring subnetwork; *p*-value, rank *p*-value calculated by comparing observed HSS with a null distribution of HSS generated for each subnetwork of a specific size via 20,000 iterations. In bold are reported the putative adaptive genes observed in multiple gene networks.

Supplementary table S3. Gene networks showing significant signatures of positive selection in the Seri population according to *signet* analysis performed on the obtained genome-wide distributions of nSL and H12 scores.

Pathway (test)	P. size	N. size	HSS	*p*-value	Genes
Starch and sucrose metabolism (nSL)	18	3	6.360	0.039	*PGM1 HK2 SI*
Ras signaling pathway (nSL)	129	3	7.401	0.011	*PRKACB CALM1 GNG2*
MAPK signaling pathway (nSL)	123	4	7.713	0.008	*CACNA1E RASGRP3 CACNA1D CACNA2D1*
Carbohydrate digestion and absorption (H12)	22	4	7.492	0.009	*ATP1A2 ATP1A4 PLCB2 CACNB4*
Calcium signaling (H12)	185	4	7.174	0.010	*ITPR2 ATP2B1 PRKG1 SLC8A1*

P. size, number of genes belonging to the pathway; N. size, number of genes composing the identified network; HSS, highest scoring subnetwork; *p*-value, rank *p*-value calculated by comparing observed HSS with a null distribution of HSS generated for each subnetwork of a specific size via 20,000 iterations. The *CALM1*, *PRKACB* and *GNG2* putative adaptive genes identified by nSl-based tests, as well as the *CACNB4*, *ITPR2*, and *PRKG1* putative adaptive loci pointed out by H12-based analyses, are known to play a role also in the *Sweet taste signaling* pathway.

Supplementary table S4. Gene networks showing significant signatures of positive selection in the Rarámuri population according to *signet* analysis performed on the obtained genome-wide distributions of nSL and H12 scores.

Pathway (test)	P. size	N. size	HSS	*p*-value	Genes
Thyroid hormone signaling pathway (nSL)	72	5	6.511	0.031	*ATP1A2 ATP1B1 THRB RCAN2 PIK3R5*
Mucin type O-Glycan biosynthesis (nSL)	25	13	6.551	0.030	*GALNT2 GALNT14 GALNT15 GALNTL6 GALNT7* ***GALNT10*** *GALNTL5 ST3GAL1 GCNT1 GALNT18 GALNT6 GALNT16 GCNT3*
Thyroid hormone signaling pathway (H12)	72	3	7.903	0.016	*MTOR RXRG STAT1*
Mucin type O-Glycan biosynthesis (H12)	25	2	6.914	0.026	***GALNT10*** *GALNT17*

P. size, number of genes belonging to the pathway; N. size, number of genes composing the identified network; HSS, highest scoring subnetwork; *p*-value, rank *p*-value calculated by comparing observed HSS with a null distribution of HSS generated for each subnetwork of a specific size via 20,000 iterations. In bold are reported the putative adaptive genes observed in multiple gene networks.

Supplementary table S5. Gene networks showing significant signatures of positive selection in populations belonging to the Central Mexican Cluster according to *signet* analysis performed on the obtained genome-wide distributions of nSL and H12 scores.

Pathway (test)	P. size	N. size	HSS	*p*-value	Genes
Glutamatergic synapse (nSL)	57	11	6.493	0.030	*PRKACB* ***GNG4*** *GNG5* ***KCNJ3*** *GRM2 GRM7* ***GNAI2*** *GRM8* ***GNG10 GNG2 GNG7***
Dopaminergic synapse (nSL)	78	8	6.965	0.017	***GNG4 KCNJ3*** *SCN1A* ***GNAI2*** *DRD2* ***GNG2 GNB5 GNG7***
Serotonergic synapse (nSL)	50	9	6.502	0.029	*GNG12* ***GNG4 KCNJ3 GNAI2 GNG10 GNG2 GNB5*** *PLCB1 PLCB4*
Chemokine signaling (H12)	162	4	5.984	0.025	***GNAI2*** *ARRB2 STAT1 RHOA*

P. size, number of genes belonging to the pathway; N. size, number of genes composing the identified network; HSS, highest scoring subnetwork; *p*-value, rank *p*-value calculated by comparing observed HSS with a null distribution of HSS generated for each subnetwork of a specific size via 20,000 iterations. In bold are reported the putative adaptive genes observed in multiple gene networks. The *ARRB2* and *GNAI2* genes belonging to the *Chemokine signaling* pathway are known to participate also in the *Dopaminergic synapse* cascade.

Supplementary table S6. Gene networks showing significant signatures of positive selection in populations belonging to the Southern Mexican Cluster according to *signet* analysis performed on the obtained genome-wide distributions of nSL and H12 scores.

Pathway (test)	P. size	N. size	HSS	*p*-value	Genes
Glutamatergic synapse (nSL)	55	5	5.831	0.040	*TRPC1 PPP3CA GRM1 PLCB1 GNAS*
Platelet activation (nSL)	67	4	6.223	0.024	*ADCY3* ***GNAI2*** *PIK3CB ADCY8*
Glutamatergic synapse (H12)	55	2	6.258	0.026	*GRIA1 PRKACB*
Chemokine signaling (H12)	162	7	5.780	0.029	***GNAI2*** *RAC1 RAC2 RHOA PAK1 ROCK1 VAV2*

P. size, number of genes belonging to the pathway; N. size, number of genes composing the identified network; HSS, highest scoring subnetwork; *p*-value, rank *p*-value calculated by comparing observed HSS with a null distribution of HSS generated for each subnetwork of a specific size via 20,000 iterations. In bold are reported the putative adaptive genes observed in multiple gene networks. The *GNAI2*, *RHOA*, and *ROCK1* putative adaptive genes belonging to the *Chemokine signaling* pathway also play a role in the *Platelet activation* cascade.

Supplementary Results

Setting the Lacandon Population Within the Fine-scale Native Mexican Genetic Structure

Consistently with results obtained by Moreno-Estrada et al. (2014), we obtained unusual F_{st} values for population pairs involving the LAC group belonging to the SMC cluster (supplementary figure S3, Supplementary Material online). In fact, LAC were previously attested to have experienced appreciable isolation, consequent reduction in population size and increased drift, which may explain the observed pattern of genetic differentiation. However, results obtained by means of the fineSTRUCTURE clustering approach, which considers also LD-related information rather than only allele frequencies, pointed to LAC as perfectly fitting within the SMC group and suggest that the observed F_{st} values might result from remarkably recent processes having affected the latest part of the history of this population (i.e. events able to alter allele frequencies, but not the length of chromosome segments shared by individuals). This finding seems to corroborate the hypothesis of recent isolation of this population in line with the long runs of homozygosity previously detected along their genomes (Moreno-Estrada et al. 2014).

Therefore, LAC are thought to have shared most of their evolutionary history with the other SMC populations and might have been thus subjected to the same long-standing selective pressures experienced by these groups. According to this picture and to the low sample size (N = 11) of the LAC population sample, despite their peculiar genetic differentiation with respect to the other studied populations (which however did not reach the level observed for SER), we thus decided to maintain this group within the SMC cluster when implementing selection scans.

Controlling for Adaptive Events Potentially Shared with Other Native American Populations

In order to test for the occurrence of adaptive events potentially shared between the considered Native Mexican clusters and other groups of Native American ancestry, we replicated selection scans and gene-network analyses on a genome-wide dataset of populations from Central and South America (Gnecchi-Ruscone et al. 2019). In particular, we focused on non-Andean South American populations belonging to the previously identified *Peruvian Amazon* genetic cluster because they were recently proposed to represent the result of migratory events from Mesoamerican ancestors into South America that are independent from those having originated the Andean populations. Moreover, multiple contacts between these South American groups and Mesoamerican populations were inferred to have occurred after the divergence between present-day Mesoamerican and South American ancestors. In fact, when tested as

forming a clade together with Andeans, these groups showed a significant extra genetic affinity with Native Mexican populations (Gnecchi-Ruscone et al. 2019), thus representing a valuable control group characterized by a largely shared genetic ancestry with Mexican people.

Selection scans and gene network analyses performed on such a South American control cluster led to the identification of six gene networks (four of them largely overlapping) that are made up of a total of 39 genes and that turned out to be significantly targeted by widespread positive selection (supplementary table S2, Supplementary Material online). Interestingly, these genes were found to differ from the putative adaptive loci identified for Native Mexican clusters (fig3, supplementary fig. S4, supplementary tables S3-S6, Supplementary Material online). Moreover, none of the significant gene networks observed for the control South American populations belongs to the functional pathways supposed to have evolved adaptively in the ancestors of present-day Mexican people (fig 3).

Validating Results from nSL-based Analyses by Computing H12 Statistic

To validate biological functions putatively targeted by widespread natural selection according to results from nSL-informed gene networks, we replicated the described gene network approach by relying on an independent selection statistic (i.e. H12). In fact, combining multiple selection scans has the potential to reduce the confounding effects due to demography or to enable the detection of different and complementary genetic footprints left by natural selection. In particular, the H12 statistic has been proved to be suitable for detecting genomic signatures ascribable to selection on standing variation (Garud et al. 2015), as is the case of nSL, which was useful to expand the range of the adaptive events investigated.

However, since nSL and H12 tests focus on different aspects of the potential adaptive haplotypes, we cannot expect to obtain exactly the same results from these analyses in terms of combinations of loci within the identified significant gene networks. Indeed, H12 is known have most power to detect events driven by strong selection (Garud et al. 2015), while nSL presents remarkable power in pinpointing signatures due to weak selection (Ferrer-Admetlla et al. 2014), which are supposed to be the most informative ones when searching for footprints of polygenic adaptation. This may explain why gene network analyses based on H12 scores pointed also to some functional pathways (e.g. *Purine metabolism, Pyrimidine metabolism*, and *Olfactory transduction*) putatively targeted by positive selection in the considered Mexican clusters that were not identified according to nSL-based analyses. These selection signatures have been observed previously in populations of East Asian ancestry (Landini et al. 2020) and might have been thus driven by strong selective events that occurred anciently in the ancestors of present-day Native American populations.

Nevertheless, H12-based results also succeeded in confirming biological functions that have been adaptively evolved according to nSL-informed gene networks and that were plausibly ascribable to more weak selective events. For instance, these signatures encompassed: i) loci involved in *Carbohydrate digestion and absorption* and *Sweet taste signaling* pathways in SER; ii) genes belonging to the *Thyroid hormone signaling* and *Mucin type O-Glycan biosynthesis* pathways in TAR; iii) genes from the *Chemokine signaling* pathway, but playing a role in the *Dopaminergic synapse* pathways in CMC; iv) loci belonging to the *Chemokine signaling* pathway, but contributing also to the *Platelet activation* cascade,in SMC (supplementary tables S3-S6, Supplementary Material online).

Supplementary References

Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R. 2014. On Detecting Incomplete Soft or Hard Selective Sweeps Using Haplotype Structure. *Mol Biol Evol* 31 (5): 1275–91. https://doi.org/10.1093/molbev/msu077.

Garud NR, Messer PW, Buzbas EO, Petrov DA. 2015. Recent selective sweeps in North American Drosophila melanogaster show signatures of soft sweeps. PLoS Genet 11 (2): e1005004. https://doi.org/10.1371/journal.pgen.1005004.

Gnecchi-Ruscone GA, Sarno S, De Fanti S, Gianvincenzo L, Giuliani C, Boattini A, Bortolini E, et al. 2019. Dissecting the Pre-Columbian Genomic Ancestry of Native Americans along the Andes-Amazonia Divide. *Mol Biol Evol* 36 (6):1254-1269. doi: 10.1093/molbev/msz066.

Landini A, Shaobo Y, Gnecchi-Ruscone GA, Abondio P, Ojeda-Granados C, Sarno S, De Fanti S, et al. 2020. Genomic Adaptations to Cereal-Based Diets Contribute to Mitigate Metabolic Risk in Some Human Populations of East Asian Ancestry. *Evol Appl* 00:1–17. Accessed September 1, 2020. https://doi.org/10.1111/eva.13090.

Lin G, Chai J, Yuan S, Mai C, Cai L, Murphy RW, Zhou W, Luo J. 2016. VennPainter: A Tool for the Comparison and Identification of Candidate Genes Based on Venn Diagrams. *PLoS One* 11 (4): e0154315. doi: 10.1371/journal.pone.0154315.

Moreno-Estrada A, Gignoux GR, Fernández-López JC, Zakharia F, Sikora M, Contreras AV, Acuña-Alonzo V, et al. 2014. The Genetics of Mexico Recapitulates Native American Substructure and Affects Biomedical Traits. *Science* 344 (6189): 1280–85. doi: 10.1126/science.1251688.