


RESEARCH

Open Access



“Like sugar in milk”: reconstructing the genetic history of the Parsi population

Gyaneshwer Chaubey^{1*†}, Qasim Ayub^{2*†}, Niraj Rai^{3,4†}, Satya Prakash³, Veena Mushrif-Tripathy⁵, Massimo Mezzavilla², Ajai Kumar Pathak^{1,6}, Rakesh Tamang⁷, Sadaf Firasat⁸, Maere Reidla^{1,6}, Monika Karmin^{1,6,9}, Deepa Selvi Rani³, Alla G. Reddy³, Jüri Parik^{1,6}, Ene Metspalu^{1,6}, Siiri Roots¹, Kurush Dalal¹⁰, Shagufta Khaliq¹¹, Syed Qasim Mehdi^{8c}, Lalji Singh¹², Mait Metspalu¹, Toomas Kivisild^{1,13}, Chris Tyler-Smith², Richard Villems^{1,6†} and Kumarasamy Thangaraj^{3*†}

Abstract

Background: The Parsis are one of the smallest religious communities in the world. To understand the population structure and demographic history of this group in detail, we analyzed Indian and Pakistani Parsi populations using high-resolution genetic variation data on autosomal and uniparental loci (Y-chromosomal and mitochondrial DNA). Additionally, we also assayed mitochondrial DNA polymorphisms among ancient Parsi DNA samples excavated from Sanjan, in present day Gujarat, the place of their original settlement in India.

Results: Among present-day populations, the Parsis are genetically closest to Iranian and the Caucasus populations rather than their South Asian neighbors. They also share the highest number of haplotypes with present-day Iranians and we estimate that the admixture of the Parsis with Indian populations occurred ~1,200 years ago. Enriched homozygosity in the Parsi reflects their recent isolation and inbreeding. We also observed 48% South-Asian-specific mitochondrial lineages among the ancient samples, which might have resulted from the assimilation of local females during the initial settlement. Finally, we show that Parsis are genetically closer to Neolithic Iranians than to modern Iranians, who have witnessed a more recent wave of admixture from the Near East.

Conclusions: Our results are consistent with the historically-recorded migration of the Parsi populations to South Asia in the 7th century and in agreement with their assimilation into the Indian sub-continent's population and cultural milieu "like sugar in milk". Moreover, in a wider context our results support a major demographic transition in West Asia due to the Islamic conquest.

Keywords: Parsi, Zoroastrian, autosomes, mtDNA, Y chromosome, ancient DNA

Background

The Parsi (or Parsee) community of the Indian sub-continent are a group of Indo-European speakers and adherents of the Zoroastrian faith, one of the earliest monotheisms that flourished in pre-Islamic Persia (present-day Iran) [1, 2]. Zoroastrianism was the religion of Persia from 600 B.C. to 650 A.D. [3–5] and, despite a

long history of well-preserved culture, it now has a limited number of followers [6, 7]. The Parzor Foundation reports the total number of Zoroastrians to be around 137,000, with 69,000 living in India, roughly 20,000 in Iran [8] and 2000–5000 in Pakistan [8, 9] (Fig. 1). This reduction in population is mainly due to strict marriage practices and low birth rates [7, 10–12].

The Parsi trace their ancestry to the ancestors of the Zoroastrians of modern Iran, who are followers of the Prophet Zoroaster or Zarathushtra [3]. In the 7th century, the Zoroastrian Sassanian dynasty was threatened by Islamic conquest and a small group of Zoroastrians fled to Gujarat in present-day India, where they were called 'Parsi' (literally meaning 'people from Paras or

* Correspondence: gyan@ebc.ee; qa1@sanger.ac.uk; thangs@ccmb.res.in

†Equal contributors

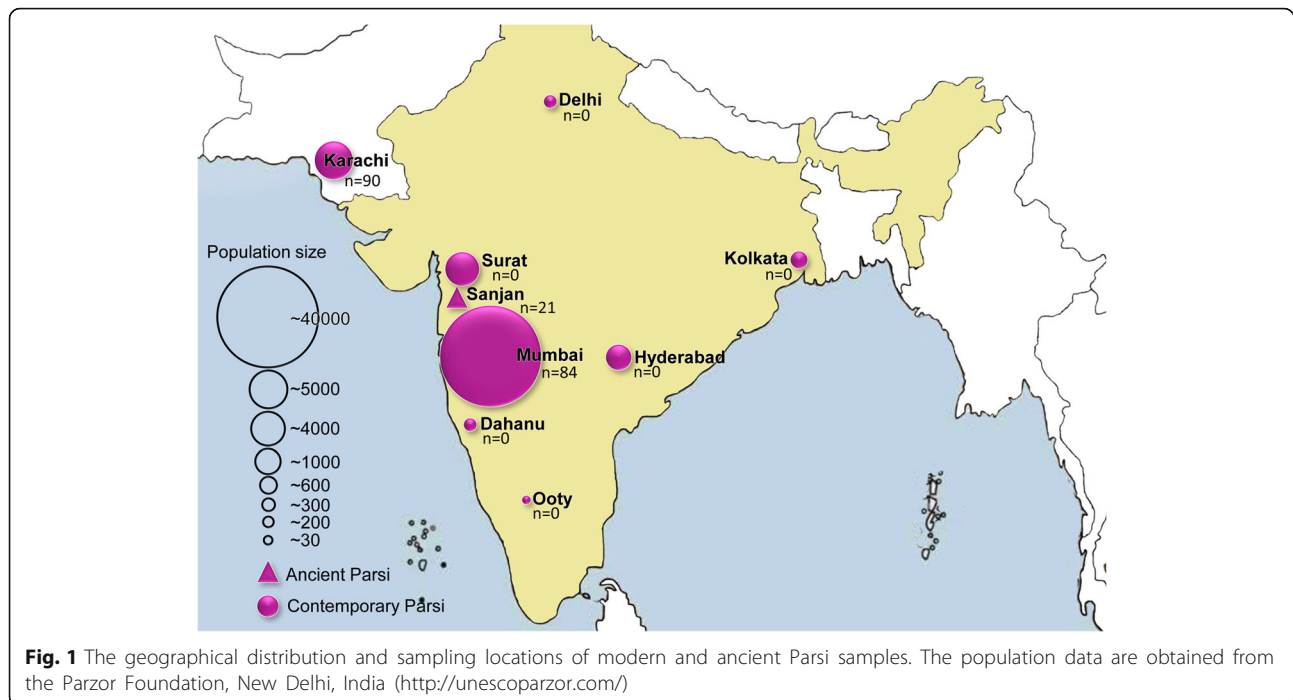
‡Deceased

¹Evolutionary Biology Group, Estonian Biocentre, Riia23b, Tartu 51010, Estonia

²The Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK

³CSIR - Centre for Cellular and Molecular Biology, Hyderabad 500007, India

Full list of author information is available at the end of the article



Fars', the local term for Persia) [3, 4, 6, 13]. Several myths narrate their first arrival in the West Coast of India and settlement in Sanjan (Gujarat) [4, 14–17]. The most popular one mentioned in the *Qissa-e-Sanjaan* is that an Indian ruler called *Jadi Rana* sent a glass full of milk to the Parsi group seeking asylum [4, 18]. His message was that his kingdom was full with local people. The Zoroastrian immigrants put sugar (or a ring, in some versions of the story) into the milk to indicate an assimilation of their people into the local society, like “sugar in milk” [14, 18]. In contemporary India and Pakistan, we see their adoption of local languages (Gujarati and Sindhi) and economic integration while maintaining their ethnic identity and practicing strict endogamy [1, 3, 4, 12, 19–21].

Previous genetic analyses of the Parsis have focussed mainly on low-resolution uni-parental markers, which have suggested their affinity with both West Eurasian and Indian populations [22–24]. Autosomal analysis based on microsatellites or human leukocyte antigens (HLA) have revealed their intermediate position among the populations of South Asia and the Middle East/Europe [9, 24]. A study of mitochondrial DNA (mtDNA) variation reported 60% of South Asia lineages among the Pakistani Parsi population [23], whereas the male lineages based on Y chromosome admixture estimates were almost exclusively Iranian [22]. Based on these results, a male-mediated migration followed by assimilation of local South Asia females was concluded [23].

These early studies of the Parsi populations relied mainly on low-resolution markers, limiting the power of

the analyses [9, 23–25], and the majority of Parsis (~98.8%), who live in India, have been underrepresented in these studies. Here we present genome-wide genotyping array data from 43 and high-resolution mtDNA and Y-chromosome genotyping data from 174 Parsi samples from India and Pakistan. In addition, we also genotyped mtDNA polymorphisms from 21 ancient Parsi samples excavated from Sanjan, in present-day Gujarat, India (Fig. 1). The human remains from Sanjan *dokhama* (tower of silence) District Valsad, Gujarat, were excavated in 2004. The accelerator mass spectrometry dating of human remains suggest that the *dokhama* belongs to the 14th to 15th century A.D. [17].

We investigated whether the current Parsi people living in India and Pakistan are genetically related amongst themselves and with the present-day Iranian population, and if their genetic composition has been affected by the neighboring Indian and Pakistani populations. We also examined runs of homozygosity (RoH) to study consanguinity. To address the extent to which the current Parsi populations assimilated local females during their long formation history, we compared their mtDNA haplogroup composition with ancient remains excavated in Sanjan, the initial settlement established by these migrants from Persia [17].

Results and discussion

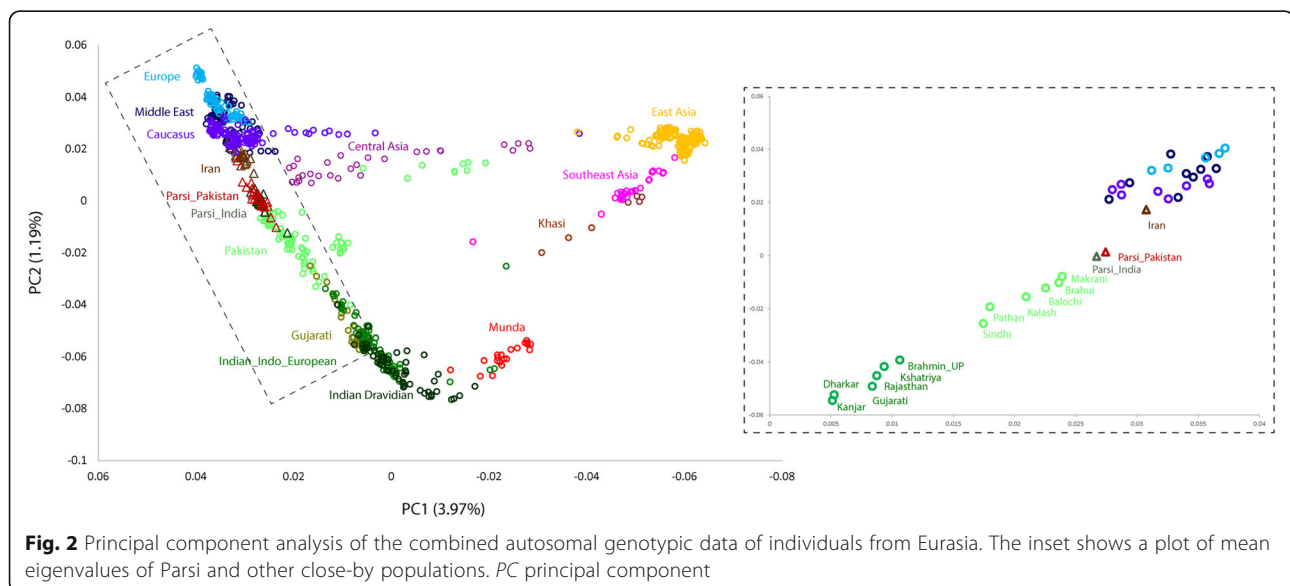
For autosomal analyses, we used Illumina HumanHap 650 K genotyping chips on 19 Indian Parsi samples collected from Mumbai, and Illumina 2.5 M genotyping chips for 24 Pakistani Parsi individuals from Karachi

(Fig. 1 and Additional file 1). The combined Parsi data set was merged with a global data set from the published literature [26–29] and references therein. Additional file 1: Table S2 lists the populations and number of single-nucleotide polymorphism (SNPs) used for various analyses after quality control (Additional file 1). The mean allele frequency differentiation between the two Parsi (Indian and Pakistani) groups was the lowest (F_{ST} Indian and Pakistani Parsi = 0.00033 ± 0.000025), followed by the differentiation of each from the Iranian population (0.011 ± 0.00021 and 0.012 ± 0.00025 for Pakistani and Indian Parsis, respectively), suggesting a common stock for both the Indian and Pakistani Parsis with the closest interpopulation affinity with populations from their putative homeland, Iran (Additional file 1: Figure S1 and Additional file 2: Table S3). Collectively, in F_{ST} -based analysis, both of the Parsi groups showed a significantly closer connection with West Eurasians than any of the Indian groups (two-tailed $P < 0.0001$).

We applied the default settings of the SMARTPCA program implemented in the EIGENSOFT package [30] and performed principal-components analysis (PCA) with other Eurasian populations using autosomal SNP data (Additional file 1: Table S2 and supplementary text). Our plot of the first and second principal components (PCs) clusters the Indian and Pakistani Parsis together, along the European–South Asian cline (Fig. 2). A plot of the population-wise mean of the eigenvalues showed their placement between the Pakistani and Iranian populations, indicating that the Parsis might have admixed from these two groups. Such an intermediate position of Parsis closer to Iranians than to their present geographic neighbors (Sindhi and Gujarati) suggests that the Parsis may have major ancestry

from West Eurasians (Iranians) and minor ancestry with South Asians (Fig. 2 and Additional file 1: Figure S1). We next applied the model-based clustering method assembled in ADMIXTURE [31] to reveal the positioning of Parsis in the genetic structure canvas of Eurasia. The best model [28, 29] suggested eight major genetic components—sometimes also referred to as “ancestral populations”—and identified the presence of three of the components within the Parsi (Fig. 3 and Additional file 1: Table S4). The distributions of these components among Indian and Pakistani Parsis were unique and resembled each other, but were distinct from their neighboring South Asian populations (Additional file 1: Figure S2a). Supporting the F_{ST} and PCA results, the Middle-Eastern-specific (blue) ancestry component was significantly higher (two-tailed $P < 0.0001$) in the Parsis than in any other populations residing in South Asia that were examined. The present-day Iranian population exhibited a striking difference from the Parsis, mainly in carrying an additional European component (light blue) and substantially lower South Asian ancestry (dark green) (Fig. 3 and Additional file 1: Figure S2a and Table S4). Furthermore, the ancestral North Indian ancestry calculated from the f_4 ancestry estimate showed a substantially higher level of this ancestry among Parsis than any other South Asia population (Additional file 1: Figure S2b).

It has been suggested previously that the Islamic conquest had a major genomic impact on several Middle Eastern populations, including Iranians [32]. Since Parsis diverged from Iranians just after this conquest, they may represent the genetic strata of Iran before the Islamic conquest. To test this scenario, we applied a formal test of admixture f_3 statistics (Additional file 1: Table S5). For Iranians, a negative value with significant Z scores



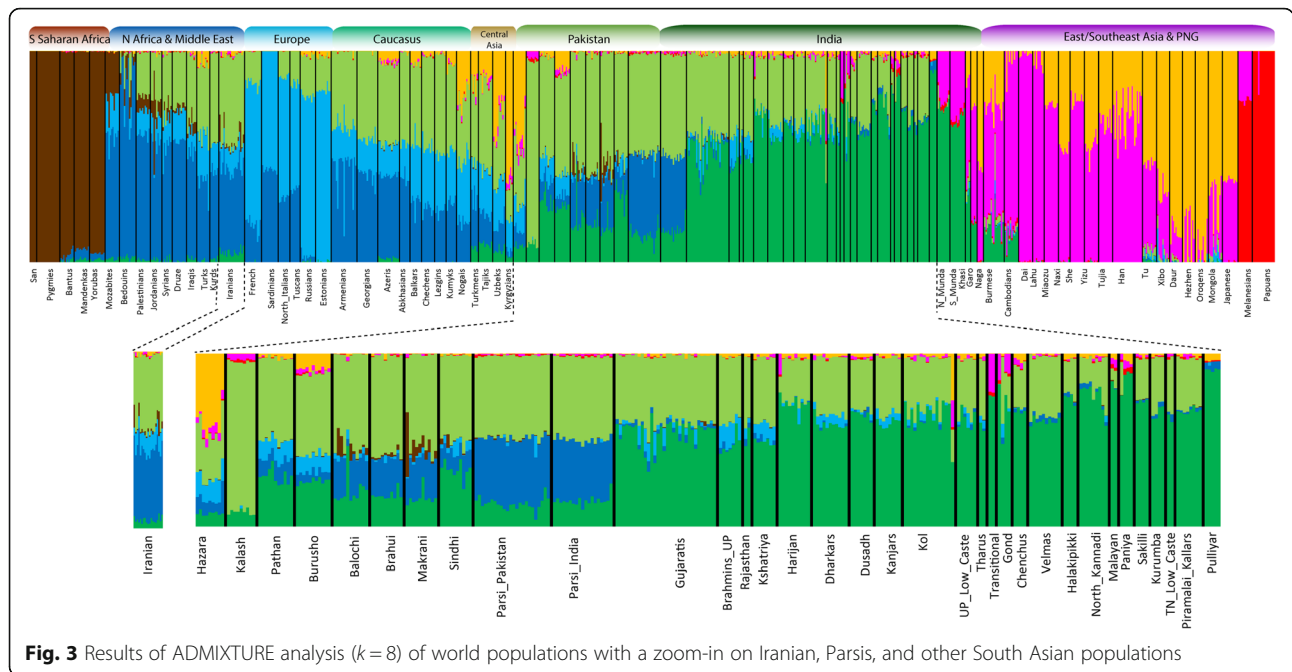


Fig. 3 Results of ADMIXTURE analysis ($k = 8$) of world populations with a zoom-in on Iranian, Parsi, and other South Asian populations

supports the hypothesis that they are descendants of a population formed by the admixture of Neolithic Iranians and populations from the Arabian Peninsula, while for Parsis this test was positive with significant Z scores. Therefore, it seems plausible that the additional light blue component we see in ADMIXTURE (Fig. 3) may have been introduced to Iran after the exile of the Parsis, likely via a recent gene flow from the Arabian Peninsula [32–34]. The quantitative estimation of South Indian and Iranian ancestries among Parsis and their South Asian neighbors showed a significant level of differentiation in ancestry composition with an inclination of Parsis towards Iranian ancestry (two-tailed $P < 0.0001$) (Table 1 and Additional file 1: Supplementary text).

We computed D statistics [35] to determine the nature of the gene flow and admixture of Parsis with their parental (Iranian) and neighboring (Gujarati and Sindhi) populations (Table 2). Consistent with the previous analyses (Figs. 2 and 3), both of the Parsi populations shared a highly significant D value with each

Table 1 The South Indian and Iranian ancestry among Parsis and neighboring populations

Population (X)	South Indian (SE)	Iranian (SE)
Pathan	31 (2.4)	56.5 (1.8)
Sindhi	22.9 (2.7)	63.1 (1.9)
Parsi (Pakistan)	6.4 (2.4)	76.6 (1.5)
Parsi (India)	8.8 (2.5)	74.6 (1.7)
Gujarati	58.1 (2)	34.3 (1.6)

South Indian ancestry: (Yoruba, Papua; X, French/Yoruba, Papua; South India, French)
 Iranian ancestry: (Yoruba, Papua; X, South India/Yoruba, Papua; Iranian, South India)

Table 2 The test of geneflow (D statistics) between Parsis, modern Iranians, Neolithic Iranians, Sindhis, and Gujaratis

Gp1	Gp2	Gp3	D value	Z score
Parsi (India)	Parsi (Pakistan)	Sindhi	0.0309	35.229
Parsi (India)	Parsi (Pakistan)	Iranians	0.0318	42.175
Parsi (India)	Gujaratis	Sindhi	-0.0015	-1.89
Parsi (Pakistan)	Parsi (India)	Sindhi	0.0309	34.184
Parsi (Pakistan)	Parsi (India)	Iranians	0.0313	39.061
Parsi (Pakistan)	Gujaratis	Sindhi	-0.002	-2.482
Sindhi	Parsi (India)	Parsi (Pakistan)	0	-0.036
Sindhi	Parsi (India)	Gujaratis	-0.0038	-4.609
Sindhi	Parsi (Pakistan)	Gujaratis	-0.0038	-4.688
Sindhi	Iranians	Parsi (Pakistan)	-0.0124	-16.085
Sindhi	Iranians	Parsi (India)	-0.0124	-15.308
Sindhi	Iranians	Gujaratis	-0.016	-17.608
Gujaratis	Parsi (Pakistan)	Parsi (India)	-0.0005	-0.898
Gujaratis	Iranians	Parsi (Pakistan)	-0.0154	-21.245
Gujaratis	Iranians	Sindhi	-0.0211	-22.669
Gujaratis	Parsi (India)	Sindhi	-0.0054	-5.982
Gujaratis	Parsi (Pakistan)	Sindhi	-0.0058	-6.7
Iran (Neolithic)	Sindhi	Iranians	-0.0002	-0.141
Iran (Neolithic)	Gujaratis	Iranians	-0.008	-5.075
Iran (Neolithic)	Parsi (India)	Iranians	0.005	3.2
Iran (Neolithic)	Parsi (Pakistan)	Iranians	0.0058	4.083

$D = (Gp1, Yoruba; Gp2, Gp3)$

other. On the other hand, the South Asian populations (Gujarati and Sindhi) had significant levels of gene flow with each other, as well as with both of the Parsi populations when evaluated with respect to the present-day Iranian population (Table 2). Two independent studies have recently reported data from ancient Iranian samples [36, 37]. It was suggested that the early Neolithic Zagros sample showed closer affinity with the Iranian Zoroastrians [36]. Here we estimated the D values of Parsis for Neolithic Iranians vs modern Iranians to compare the allele sharing. Our results demonstrated a significant level of genetic affinity between Parsis and Neolithic Iranians (Table 2 and Additional file 1: supplementary text and Table S6). The outgroup f_3 statistics of ancient Iranian samples supported the close affinity of the Parsis with Neolithic Iranians (Additional file 1: Supplementary text, Figure S3, and Table S6). Moreover, for modern populations, the outgroup f_3 statistic test and identity-by-state plots supported the closer affinity of the Parsis with the West Eurasian populations than South Asians (Additional file 1: Figure S4 and S5). To compare the shared drift with Iranian and Indian (South Munda) populations (Additional file 1: Figure S6; see Additional file 1: supplementary text for justification of the use of South Munda to represent Indian ancestry), we plotted the derived allele sharing values of Parsis and other Eurasian populations calculated with respect to the Iranian and South Munda (Indian) populations (Additional file 1: supplementary text and Figure S6). This analysis aligned the Parsis closer to the Iranian axis between Pakistani and West Eurasian populations, supporting the historical interpretation of the most recent common ancestry of Parsis with the Iranians. A TreeMix [38] analysis supports these conclusions and shows the Parsis located between the South Asian and Iranians (Additional file 1: Figure S7).

We computed a maximum likelihood tree and co-ancestry matrix based on the haplotype structure of the Parsi populations, applying the default settings of ChromoPainter and fineSTRUCTURE (version 1) [39]. The maximum likelihood tree split South Asian and West Eurasian populations into two distinct clusters (Additional file 1: Figure S8). All the Parsi individuals form a unique sub-cluster embedded within the major West Eurasian population trunk. The co-ancestry matrix plot clearly differentiated Parsis from their neighbors in

sharing a large number of chunk counts with West Eurasian (mainly Iranian and Middle Eastern) populations (Additional file 1: Figure S9 and supplementary text). Additionally, South Asian populations have donated a significantly higher number of chunks to Parsis than they received from them (two tailed $P < 0.0001$). However, the number of these chunks were significantly lower than the chunk counts shared between any pair of South Asian populations (two-tailed $P < 0.0001$) (Additional file 1: Figure S9). The fineSTRUCTURE and D statistic results thus largely suggest unidirectional minor gene flow from South Asians to Parsis (Table 2 and Additional file 1: Figure S9).

We used ALDER, a method based on linkage disequilibrium [40], to estimate the time of admixture between Parsis and their neighboring South Asian populations. For this analysis, we used the present-day Iranian vs Gujarati or Sindhi populations as their surrogate ancestors. We estimated the admixture time of Parsi groups to be around ~ 40 generations (95% CI 26–50), which yields a time of 1160 years (assuming a generation time of 29 years), in good agreement with their historically recorded migration to South Asia (Table 3). We also tested evidence for a more complex admixture history using MALDER [41], which can be used to infer multiple admixture events. The MALDER analyses confirmed the ALDER results, demonstrating only one admixture event 54 ± 8 generations ago. The ancestral sources with the highest amplitude in MALDER were Sardinian and Dai (Additional file 1: Table S7).

To investigate further the parental relatedness among Parsis [19], we analyzed the RoH and inbreeding coefficient in the population (Additional file 1: Figure S10). For RoH calculations, we applied three window sizes (1000, 2500, and 5000 kb), requiring a minimum of 100 SNPs per window and allowing one heterozygous and five missing calls per window. Long RoH segments characterize consanguinity and also provide a distinctive record of the demographic history for a particular population [42, 43]. As expected, both of the Parsi populations carried a larger number of long segments relative to their putative parental populations and present neighbors at the 1000-kb window length, likely due to the small population size and a high level of inbreeding. However, the Sindhi population from Pakistan also showed a higher level of inbreeding at the larger RoH

Table 3 The formal text of admixture using the ALDER method

Reference 1	Reference 2	Admixed	Generation time	P value	Z score
Iranian	Gujarati	Parsi (India)	38.26 ± 12.16	0.0017	3.15
Iranian	Sindhi	Parsi (India)	32.96 ± 9.42	0.013	2.48
Iranian	Sindhi	Parsi (Pakistan)	41.32 ± 8.93	1.7×10^{-5}	4.3
Iranian	Gujarati	Parsi (Pakistan)	30.74 ± 14.04	0.029	2.19

window sizes, most likely due to an elevated level of cross-cousin marriages (Additional file 1: Figure S10).

To investigate how random genetic drift has shaped the functional genetic variation after admixture in the Parsis, we implemented the population branch statistic [44] using the Sindhi and Iranians as reference and outgroup. We analysed variants over the 99.9th percentile of the genomic distribution, focussing only on those that were annotated as missense, stop gain, stop loss, splice acceptor, or splice donor using the Ensembl Variant Effect Predictor tool [45]. This revealed a cluster of linked SNPs in the HLA region and a missense SNP in CD86 (rs1129055) with a high ancestral G allele frequency in the Parsi (0.87) (Additional file 1: Figure S11 and Table S8). The frequency of this G allele is lower in the Iranians and Sindhi (0.60) and other South Asians (0.52) and East Asians (0.40). This polymorphism has been associated with the pathogenesis of pneumonia-induced sepsis and the G allele has been associated with a decreased risk of active brucellosis in Iranians [46]. The G variant has also been associated as an eQTL for decreased expression of *IQBC1*, an IQ-motif-containing B1 gene that is highly expressed in Epstein–Barr virus-transformed B lymphocytes [47].

To obtain a detailed understanding of the sex-specific South Asian and Iranian ancestries, we examined maternally inherited mtDNA and paternally inherited Y chromosome biallelic polymorphisms in a larger sample in both the Indian and Pakistani Parsi populations (Additional file 1: Figure S12, Tables S9 and S10). For the mtDNA analysis, we were also able to assay 21 ancient samples from the Sanjan [21] region, for 108 diagnostic polymorphisms (Additional file 1: Table S11 and supplementary text). Interestingly, we observed 48% South-Asian-specific lineages (haplogroups M2, M3, M5, and R5) among the ancient Parsi samples, which could potentially be explained in two ways. First, these haplogroups might have been carried by the migration of Zoroastrian refugees from Fars (Iran), a possibility that is supported by the presence of these clades in present-day Persian samples (9.9%) [34]. Second, they might have resulted from the assimilation of local females during the initial settlement. The comparison of ancient and modern samples thus identified maternal lineages that can be considered as founding (surviving or lost), as well as those that were subsequently assimilated (Additional file 1: Figure S12 and Tables S9–11). The Y chromosome profiles of Indian and Pakistani Parsi populations revealed a higher frequency of Middle-Eastern-specific lineages than South Asian ones in the Parsis (Additional file 1: Figure S12 and Table S10). The PC analysis of both mtDNA and Y chromosome data placed all the Parsi groups close to each other and showed their contrasting clustering based on maternal or paternal

ancestries (Fig. 4). For mtDNA, the Parsi cluster was closer to the Indian and Pakistani cluster (Fig. 4a), whereas for the Y chromosome it aligns between the Iranian and Pakistani populations (Fig. 4b). The Y-chromosomal PCA is similar to the autosomal PCA (Fig. 2 and Additional file: Figure S1). The contrasting patterns of maternal and paternal ancestry support a largely female-biased admixture from the South Asian populations to the Parsis.

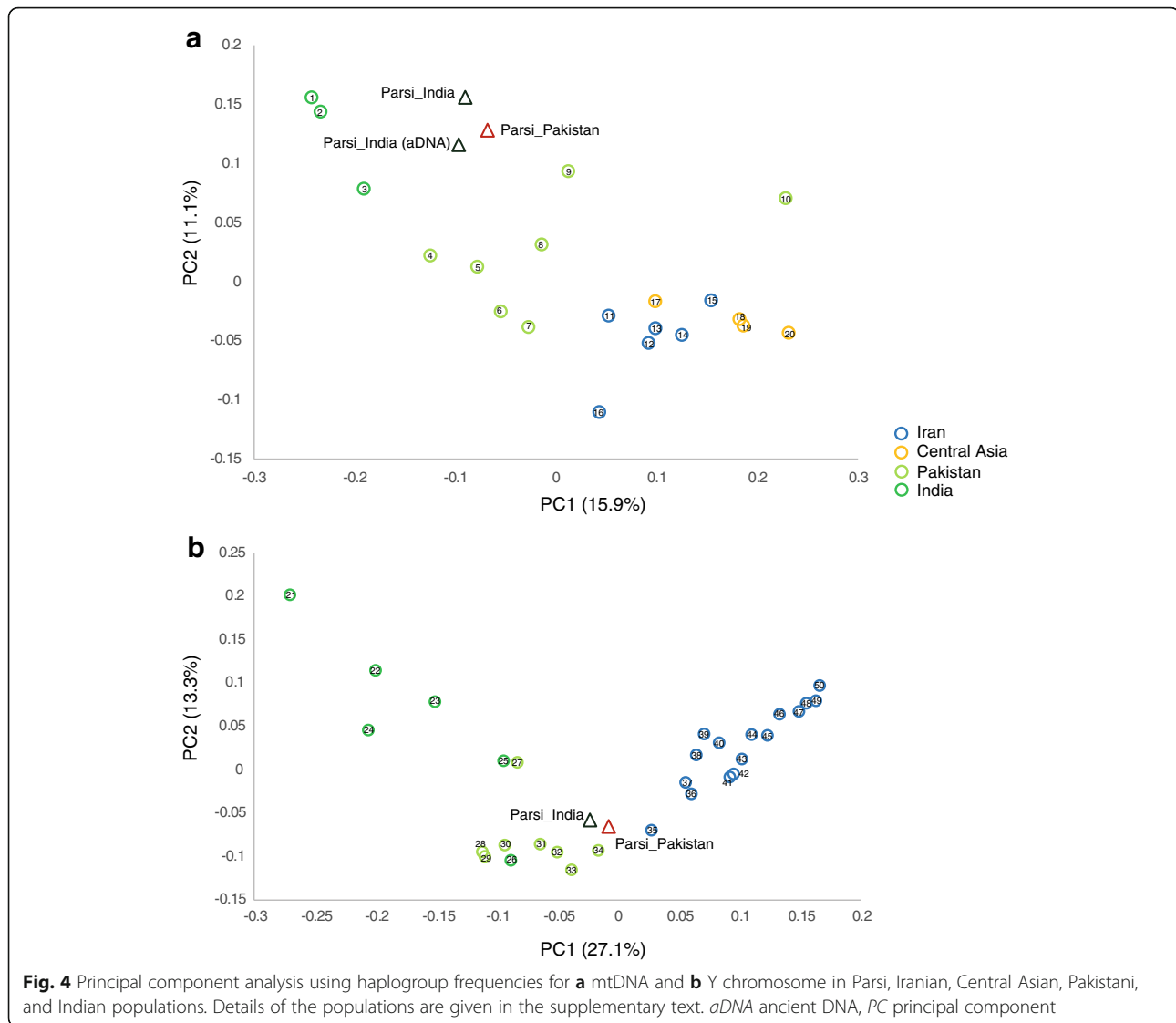
Conclusions

In conclusion, our investigation has not only contributed substantial new data, but has also provided a more comprehensive insight into the population structure of Parsis and their genetic links to Iranians and South Asians. We show that the Parsis are genetically closer to Iranian and Caucasian populations than those in South Asia and provide evidence of sex-specific admixture with the prevailing female gene flow from South Asians to the Parsis. Our results are consistent with the historically recorded migration of the Parsi populations to South Asia in the 7th century and in agreement with their assimilation into the Indian sub-continent's population and cultural milieu “like sugar in milk”.

Methods

A detailed description of the material and methods can be found in the supplementary text (Additional file 1). The modern Parsi samples were pooled from three independent collections: two from Mumbai, India, and one from Karachi, Pakistan (Fig. 1). Illumina 650 K and 2.5 M chips were used to genotype 19 Indian and 24 Pakistani Parsi individuals, respectively, following the manufacturer's specifications. We merged our newly generated data of 43 samples with the relevant reference data sets of 829 samples published elsewhere (Additional file 1: supplementary text and Table S2). For mtDNA control and coding region polymorphisms, we genotyped 117 Indian and 50 Pakistani Parsi samples (Additional file 1: Table S9). We followed phylotree (build 17) to classify them into haplogroups. For Y chromosome genotyping, 90 Pakistani samples were genotyped either by sequencing or by (PCR-RFLP) Polymerase Chain Reaction- Restriction Fragment Length Polymorphism for the relevant Y chromosome markers, whereas 84 Indian samples were assayed for 80 Y chromosomal SNPs using Sequenom mass array technology (Additional file 1: Table S10).

Ancient DNA samples were excavated from Sanjan, Gujarat, in 2001 (Additional file 1: supplementary text). Archaeological analysis and accelerator mass spectrometry dating were consistent with these remains belonging most likely to migrant Parsis from the 8–13th centuries (Additional file 1: supplementary text). The teeth obtained from 21 of these specimens were analyzed at the ancient DNA laboratory of the Centre for Cellular



and Molecular Biology of the Council of Scientific and Industrial Research (CSIR), Hyderabad, India. We followed our standard published protocol to isolate DNA from teeth [48] (Additional file 1: supplementary text).

For autosomal analyses, after data curation and merging (Additional file 1: supplementary text), we first used the method of Cockerham and Weir [49] to estimate the mean pairwise F_{ST} . Further, we performed PCA on pruned data using SMARTPCA v.7521 [30] (with default settings). We also used the F_{ST} :Yes method of SMARTPCA to calculate the F_{ST} with standard errors. We ran unsupervised ADMIXTURE v.1.23 for 25 times for each $K = 2$ to $K = 12$, and used the method described previously to choose the best K value [28–29]. The F statistics were calculated by the ADMIXTOOLS package v.3 [35] and the haplotype-based analysis was performed by Chromainter and fineSTRUCTURE v.1 [39]. The maximum likelihood tree of world populations was constructed using

TreeMix v.1.12 [38] and the RoH were calculated using PLINK 1.9 [50]. ALDER v 1.03 [40] and MALDER v.1.0 [41] were used to calculate the time and number of admixture events. The population branch statistic method [44] was used to identify genomic regions under selection in the Parsi population.

Additional files

Additional file 1: Supplementary text explaining the archeological details of ancient samples; isolation of ancient DNA, genotyping, statistical analyses and peopling of South Asia and Parsi chapters. 12 figures and 10 tables are also incorporated in this file. (PDF 13264 kb)

Additional file 2: Table S3. The population-wise F_{ST} values based on an autosomal data set for all the populations included in this study. (XLSX 120 kb)

Acknowledgements

We are grateful to the Parsi community of India and Pakistan for donating their samples. We are also thankful to Dr. Shernaz Cama, Director, Parzor

Foundation, New Delhi, India, for her help and critical comments and Tuuli Reisberg for technical assistance. GC thanks Giacomo Benedetti and Alberto Gonzalez for a useful discussion. The analyses were performed in the High Performance Computer Center of the University of Tartu, Estonia, and the Wellcome Trust Sanger Institute, Hinxton, UK. We dedicate this paper to the memory of our esteemed colleague, SQM. May his soul rest in peace.

Funding

Support was provided by Estonian personal grants PUT-766 (GC, MK, and AKP); the EU European Regional Development Fund through the Centre of Excellence in Genomics to the Estonian Biocentre and project 2014–2020.4.01.15-0012, and Estonian Institutional Research grants IUT24-1 (RV, MM, SR, EM, MR, and TK); CSIR, Government of India (KT); Wellcome Trust grant 098051 (QA and CTS); World Zarathushti Cultural Foundation, Parsi Foundation, and the Indian Archaeological Society (VMT); PIRSES-GA-2012-318979 grant (MK) DST, Government of India (LS), and an ERC Starting Investigator grant (FP7 - 261213) (TK). AKP was supported by the European Social Fund's Doctoral Studies and Internationalisation Programme DoRa. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

The data are available from the Gene Expression Omnibus of the National Center for Biotechnology Information (accession GSE97086) and the data repository of the Estonian Biocentre (www.ebc.ee/free_data).

Authors' contributions

GC, QA, RV, KT, and NR conceived the project and designed the experiments. GC, QA, NR, SP, AKP, RT, SF, MR, MK, DSR, AGR, JP, EM, SR, and SK carried out haploid DNA genotyping and analyses. GC, QA, MMz, and TK analysed the autosomal data. VMT and KD performed the archaeological study. VMT and NR collected ancient samples. NR isolated DNA from ancient remains and performed genotyping. SQM, LS, MM, CTS, RV, and KT provided samples and reagents. GC, QA, NR, VMT, and MMz wrote the manuscript with the help of TK, CTS, RV, MM, and KT. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Ethics approval and consent to participate

Informed written consent was obtained from all the participants. The ethics committees of the participating institutions all approved the study: Research Ethics Committee of the University of Tartu (approval 252/M-17); Institutional Ethics Committee, CSIR Centre for Cellular and Molecular Biology, Hyderabad, India; Human Materials and Data Management Committee, Wellcome Trust Sanger Institute, UK; and Human Subjects Committee at the Biomedical & Genetic Engineering Division, Islamabad, Pakistan. All the experimental methods comply with the Helsinki Declaration.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Evolutionary Biology Group, Estonian Biocentre, Riia23b, Tartu 51010, Estonia. ²The Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK. ³CSIR - Centre for Cellular and Molecular Biology, Hyderabad 500007, India. ⁴Present address: Birbal Sahni Institute of Palaeosciences, Lucknow 226007, India. ⁵Department of Archaeology, Deccan College Post-Graduate and Research Institute, Pune, Maharashtra 411006, India. ⁶Department of Evolutionary Biology, Institute of Molecular and Cell Biology, University of Tartu, Tartu 51010, Estonia. ⁷Department of Zoology, University of Calcutta, Kolkata 700073, India. ⁸Centre for Human Genetics and Molecular Medicine, Sindh Institute of Urology and Transplantation, Karachi 74200, Pakistan. ⁹Department of Psychology, University of Auckland, Auckland 1142, New Zealand. ¹⁰Centre for Archaeology (CfA), Centre for Extra Mural Studies (CEMS) University of Mumbai (Kalina Campus) Vidyanaagri, Santacruz E Mumbai 400098, India. ¹¹Department of Human Genetics & Molecular Biology, University of Health Sciences, Lahore 54000, Pakistan. ¹²Genome foundation, C/o Prasad Hospital, Nacharam, Hyderabad 500076, India. ¹³Division of Biological Anthropology, University of Cambridge, Cambridge CB2 3QG, UK.

Received: 1 February 2017 Accepted: 23 May 2017

Published online: 14 June 2017

References

1. The Parsis of India: Preservation of Identity in Bombay City. Leiden: Brill Academic Publishers. 2001;368.
2. Axelrod P. Myth and identity in the Indian Zoroastrian community. *J Mithraic Stud.* 1980;3(1–2):150–65.
3. Mirza HDK. Outlines of Parsi history. Mumbai: Mirza; 1974.
4. Hinnells J, Williams A. Parsis in India and the Diaspora. Routledge. 2007;304.
5. Hinnells JR. The modern Zoroastrian diaspora. In: Brown JM, Foot R, editors. Migration: the Asian experience. New York: St Martin's Press; 1994. p. 56–82.
6. Roy TK, Unisa S, Bhatt M. Growth of Parsi population in India. Mumbai: National Commission for Minorities; 2004.
7. Katrak SK. Who are the Parsees? Karachi: Pakistan Herald Press; 1965.
8. Parzor Foundation, New Delhi. <http://unescoparzor.com>. Accessed 18 May 2017.
9. Mohyuddin A, Mehdi S. HLA analysis of the Parsi (Zoroastrian) population in Pakistan. *Tissue Antigens.* 2005;66(6):691–5.
10. Pocha JS. Parsis: the vanishing breed. London: Archaeological and Cultural News of the Iranian World, CAIS; 2007.
11. Modi B. Parsi Gujarati: vanishing dialect: vanishing culture. München, Lincom Europa; 2011.
12. Shroff Z, Castro MC. The potential impact of intermarriage on the population decline of the Parsis of Mumbai, India. *Demogr Res.* 2011;25:545–64.
13. Sekar CC. Some aspects of Parsi demography. *Hum Biol.* 1948;20(2):47–89.
14. Paymaster RB. Early history of the Parsees in India from their landing in Sanjan to 1700 AD. Zartoshti Dharam Sambandhi Kelavni Apnari ane Dnyan Felavnari Mandli. 1954.
15. Gupta S, Dalal KE, Dandekar A, Nanji R, Aravazhi P, Bomble S, et al. On the footsteps of Zoroastrian Parsis in India excavations of Sanjan on the West Coast. 2003;1:93–6.
16. Gupta S, Dalal K, Dandekar A, Nanji R, Mitra R, Pandey R, et al. A preliminary report on the excavations at Sanjan (2002). *Puratattva.* 2002;32:182–98.
17. Mushrif-Tripathy V, Walimbe SR. Human skeletal remains from the medieval site of Sanjan: osteobiographic analysis. *Archaeopress*; 2012
18. Singh N. The sugar in the milk: the Parsis in India. India: Institute for Development Education (Madras, India), Madras Parsi Zarthosti Anjuman. 1986. p. 149.
19. Karkal M. Marriage among Parsis. *Demography India.* 1975;4(1):128–45.
20. Karkal M. Survey of Parsi population of greater Bombay. Bombay: Bombay Parsi Panchayat; 1982.
21. Jonnalagadda M, Ozarkar S, Mushrif-Tripathy V. Population affinities of Parsis in the Indian subcontinent. *Int J Osteoarchaeol.* 2011;21(1):103–10.
22. Qamar R, Ayub Q, Mohyuddin A, Helgason A, Mazhar K, Mansoor A, et al. Y-chromosomal DNA variation in Pakistan. *Am J Hum Genet.* 2002;70(5):1107–24.
23. Quintana-Murci L, Chaix R, Wells RS, Behar DM, Sayar H, Scozzari R, et al. Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor. *Am J Hum Genet.* 2004;74(5):827–45.
24. Rosenberg NA, Mahajan S, Gonzalez-Quevedo C, Blum MG, Nino-Rosales L, Niniis V, et al. Low levels of genetic divergence across geographically and linguistically diverse populations from India. *PLoS Genet.* 2006;2(12):e215.
25. Undevia JV. Population genetics of the Parsis. Florida: Field Museum; 1972.
26. Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, et al. Worldwide human relationships inferred from genome-wide patterns of variation. *Science.* 2008;319:1100–4.
27. Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke I, et al. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature.* 2014;505(7481):87–91.
28. Chaubey G, Metspalu M, Choi Y, Mägi R, Romero IG, Soares P, et al. Population genetic structure in Indian Austroasiatic speakers: the role of landscape barriers and sex-specific admixture. *Mol Biol Evol.* 2011;28(2):1013–24.
29. Metspalu M, Romero IG, Yunusbayev B, Chaubey G, Mallick CB, Hudjashov G, et al. Shared and unique components of human population structure and genome-wide signals of positive selection in South Asia. *Am J Hum Genet.* 2011;89(6):731–44.
30. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet.* 2006;2(12):e190.
31. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 2009;19(9):1655–64.

32. Haber M, Gauguier D, Youhanna S, Patterson N, Moorjani P, Botigué LR, et al. Genome-wide diversity in the Levant reveals recent structuring by culture. *PLoS Genet.* 2013;9(2):e1003316.
33. Grugni V, Battaglia V, Hooshiar Kashani B, Parolo S, Al-Zahery N, Achilli A, et al. Ancient migratory events in the Middle East: new clues from the Y-chromosome variation of modern Iranians. *PLoS One.* 2012;7(7):e41252.
34. Derenko M, Malyarchuk B, Bahmanimehr A, Denisova G, Perkova M, Farjadian S, et al. Complete mitochondrial DNA diversity in Iranians. *PLoS One.* 2013;8(11):e80673.
35. Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, et al. Ancient admixture in human history. *Genetics.* 2012;192(3):1065–93.
36. Broushaki F, Thomas MG, Link V, López S, van Dorp L, Kirsanow K, et al. Early Neolithic genomes from the eastern Fertile Crescent. *Science.* 2016;353:499–503.
37. Lazaridis I, Nadel D, Rollefson G, Merrett DC, Rohland N, Mallick S, et al. Genomic insights into the origin of farming in the ancient Near East. *Nature.* 2016;536:419–24.
38. Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 2012;8(11):e1002967.
39. Lawson DJ, Hellenthal G, Myers S, Falush D. Inference of population structure using dense haplotype data. *PLoS Genet.* 2012;8(1):e1002453.
40. Loh PR, Lipson M, Patterson N, Moorjani P, Pickrell JK, Reich D, et al. Inferring admixture histories of human populations using linkage disequilibrium. *Genetics.* 2013;193(4):1233–54.
41. Pickrell JK, Patterson N, Loh PR, Lipson M, Berger B, Stoneking M, et al. Ancient west Eurasian ancestry in southern and eastern Africa. *Proc Natl Acad Sci U S A.* 2014;111(7):2632–7.
42. McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, et al. Runs of homozygosity in European populations. *Am J Hum Genet.* 2008;83(3):359–72.
43. Kirin M, McQuillan R, Franklin CS, Campbell H, McKeigue PM, Wilson JF, et al. Genomic runs of homozygosity record population history and consanguinity. *PLoS One.* 2010;5(11):e13996.
44. Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, Pool JE, et al. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science.* 2010;329(5987):75–8.
45. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F, et al. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics.* 2010;26(16):2069–70.
46. Eskandari-Nasab E, Moghadampour M, Najibi H, Hadadi-Fishani M. Investigation of CTLA-4 and CD86 gene polymorphisms in Iranian patients with brucellosis infection. *Microbiol Immunol.* 2014;58(2):135–41.
47. The Genotype-Tissue Expression (GTEx) pilot analysis. Multitissue gene regulation in humans. *Science.* 2015;348(6235):648.
48. Rai N, Taher N, Singh M, Chaubey G, Jha AN, Singh L, et al. Relic excavated in western India is probably of Georgian Queen Ketevan. *Mitochondrion.* 2014;14(1):1–6.
49. Cockerham CC, Weir BS. Covariances of relatives stemming from a population undergoing mixed self and random mating. *Biometrics.* 1984;40(1):157–64.
50. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ, et al. Second-generation PLINK: rising to the challenge of larger and richer data sets. *BMC Biol.* 2015;4(1):1–16.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

