Deciphering the Signature of Selective Constraints on Cancerous Mitochondrial Genome

Jia Liu,†,[1,2,3] Li-Dong Wang,†,[4] Yan-Bo Sun,†,[1,2,3] En-Min Li,[5] Li-Yan Xu,[5] Ya-Ping Zhang,[1,2,6] Yong-Gang Yao,[7] and Qing-Peng Kong*,[1,2]

[1]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China

[2]KIZ/CUHK Joint Laboratory of Bioresources and Molecular Research in Common Diseases, Kunming, China

[3]Graduate School of the Chinese Academy of Sciences, Beijing, China

[4]Henan Key Laboratory for Esophageal Cancer Research, Department of Basic Oncology and Pathology at College of Medicine, The First and The Second Affiliated Hospital, Zhengzhou University, Zhengzhou, China

[5]Department of Biochemistry, Medical College of Shantou University, Shantou, China

[6]Laboratory for Conservation and Utilization of Bio-resources, Yunnan University, Kunming, China

[7]Key Laboratory of Animal Models and Human Disease Mechanisms, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China

†These authors contributed equally to this work.

***Corresponding author:** E-mail: kongqp@yahoo.com.cn.

Associate editor: David Irwin

Abstract

In accordance with the hypothesis that cancer formation is a process of somatic evolution driven by natural selection, signature of positive selection has been detected on a number of cancer-related nuclear genes. It remains, however, controversial whether a similar selective pressure has also acted on mitochondrial DNA (mtDNA), a small molecule in mitochondrion that may play an important role in tumorigenesis by altering oxidative phosphorylation. To better understand the mutational pattern on cancerous mtDNA and decipher the genetic signature left by natural selection, a total of 186 entire mitochondrial genomes of cancerous and adjacent normal tissues from 93 esophageal cancer patients were obtained and extensively studied. Our results revealed that the observed mutational pattern on the cancerous mtDNAs might be best explained as relaxation of negative selection. Taking into account an additional 1,235 cancerous (nearly) complete mtDNA sequences retrieved from the literature, our results suggested that the relaxed selective pressure was the most likely explanation for the accumulation of mtDNA variation in different types of cancer. This notion is in good agreement with the observation that aerobic glycolysis, instead of mitochondrial respiration, plays the key role in generating energy in cancer cells. Furthermore, our study provided solid evidence demonstrating that problems in some of the published cancerous mtDNA data adequately explained the previously contradictory conclusions about the selective pressure on cancer mtDNA, thus serving as a paradigm emphasizing the importance of data quality in affecting our understanding on the role of mtDNA in tumorigenesis.

Key words: cancer, mitochondrial genome, somatic mutation, selective constraints.

Introduction

The formation of cancer was recognized as a process of somatic evolution driven by natural selection (Nowell 1976), in which any genetic alterations that confer selective advantage to the fitness of the malignant cell would then be preferentially kept. In accordance with this hypothesis, signature of positive selection has been detected on a number of cancer-related genes, opening a new avenue in understanding the molecular mechanism underling carcinogenesis (Crespi and Summers 2006).

Although it is known that positive selection has played an important role in shaping the genetic landscape of nuclear genome in cancer cells, it remains controversial whether a similar selective pressure has also acted on mitochondrial DNA (mtDNA), which is located in mitochondrion but suggested to play an important role in tumorigenesis by altering mitochondrial oxidative phosphorylation (cf. Chandra and Singh 2011). In the past decade, a high frequency of somatic mtDNA mutations have been observed in a variety of cancers and were regarded as solid evidence supporting an important role of mtDNA during the neoplastic process (cf. Wallace 2005; Brandon et al. 2006; Chatterjee et al. 2006). Consistent with this notion, a recent study suggested that cancerous mitochondrial genomes were likely under positive selection (Zhidkov et al. 2009), further suggesting that mtDNA mutations likely have some functional potential in carcinogenesis. Although it has been proposed that random process can fully explain a large fraction of the observed somatic mtDNA mutations in cancer (Coller et al. 2001), a higher ratio of nonsynonymous to synonymous substitutions among these mutations indicated the existence of

selection (Stafford and Chen-Quin 2010). Contrasting with the scenario proposed by Zhidkov et al. (2009), Stafford and Chen-Quin (2010) suggested that the mitochondrial genomes in cancer cells were most plausibly shaped by the relaxed selective constraints based on the observation that some mtDNA genes in cancer cells had a higher missense mutation rate than that in the general population. Evidence in support of the latter viewpoint came from the observation that aerobic glycolysis, instead of mitochondrial respiration, plays a key role in generating energy in cancer cells (Vander Heiden et al. 2009). Although both studies have obtained their findings by reanalyzing the mitochondrial genome data retrieved from the literature, their opinions about the selective pressure on cancerous mtDNA were surprisingly contradictory (Zhidkov et al. 2009; Stafford and Chen-Quin 2010). These conflicting conclusions were then likely introduced by problems in data quality, as had been discussed previously (Salas et al. 2005). Scrutinizing the mtDNA data from Zhou et al. (2007), which were considered by both studies, has revealed a number of suspicious somatic mutations. (supplementary table S1, Supplementary Material online). For instance, the observed "somatic mutations" in head and neck cancerous tissue of patient 1836 consist of virtually all diagnostic variants (namely 499, 1811, 4646, 5999, 6047, 7705, 11332, 11467, 12308, 12372, 14620, 15693, 16278, 16356) of haplogroup U4b (van Oven and Kayser 2009); since it is unlikely that these somatic mutational events occurred "coincidently" on all the U4b diagnostic sites, the most reasonable explanation would be that the cancerous tissue of the patient has been suffered from sample mix-up or contamination. Alternatively, it is also possible that signature of selective pressure may vary across different types of cancer and thus explains the conflicting opinions between both studies (Zhidkov et al. 2009; Stafford and Chen-Quin 2010). More mitochondrial genome data from additional cancer would then help to clarify these issues.

To better understand the mutational pattern on cancerous mtDNA and decipher the genetic signature left by natural selection, a total of 186 entire mitochondrial genomes of the cancerous and adjacent normal tissues from 93 esophageal cancer patients were obtained in this study under a stringent quality control procedure (see "Sample collection and sequencing" in Materials and Methods). Based on this new data set, supplemented by our previously reported 90 mitochondrial genomes from Chinese patients with breast cancer (Wang et al. 2007) or colorectal cancer (Wang et al. 2011), our results revealed that mtDNA in esophageal cancer is most likely driven by the relaxed negative selection, which is a common hallmark among different types of cancer. Significantly, our study provides solid evidence disclosing that problems in some of the published cancerous mtDNA data sets introduced by sample mix-up or contamination of exogenous DNA well explain the previously conflicting viewpoints about the selective pressure on cancer mtDNA, thus serving as a paradigmatic example emphasizing the importance of data quality in affecting our understanding on the role of mtDNA in tumorigenesis.

Materials and Methods

Sample Collection and Sequencing

In total, 93 surgically resected esophageal cancer specimens were obtained from Taihang area in Henan province, China, a high-incidence area for esophageal cancer. None of the patients received any radiotherapy or chemotherapy before surgery. Histopathologically, all the patients had been confirmed with esophageal squamous cell carcinoma. For each patient, both cancerous and adjacent normal tissues were collected and stored under −80 °C until DNA extraction. Genomic DNA from each issue was isolated by using the phenol/chloroform extraction method. To avoid the problems of artificial recombination and mitochondrial pseudogenes (Salas et al. 2005; Kong et al. 2008; Yao et al. 2008), the whole mtDNA genome was amplified into two overlapping polymerase chain reaction fragments, each comprising about 8,500 bp in length (Fendt et al. 2009). Sequencing was performed by BigDye Terminator v3.1 Cycle Sequencing Kit and run on 3730 DNA Analyzer (Applied Biosystems). Sequences were aligned and edited by the LaserGene v7.1 software (DNAStar, Inc., Madison, WI), and mutations were recorded by comparing with the revised Cambridge reference sequence (rCRS; Andrews et al. 1999). To further ensure the quality of the obtained sequence data, a stringent quality control procedure (Kong et al. 2003, 2006; Palanichamy et al. 2004; Sun et al. 2006) and some caveats (Kong et al. 2008; Bandelt et al. 2009; Yao et al. 2009) were adopted during the course of sample handling and data generation. Specifically, every sample was sequenced at least twice, and all the recorded mutations were confirmed by rechecking the sequencing electropherograms and/or resequencing. Moreover, for any observed somatic mutation(s) in certain tissue, the mutation(s) was further validated by reamplifying and sequencing the genomic DNA of the same tissue sample.

Evaluating the Collected Cancerous Mitochondrial Genome Data

Besides the newly generated complete mtDNAs, we also analyzed the mtDNA data studied by Zhidkov et al. (2009) and Stafford and Chen-Quin (2010). The original and "corrected" mtDNA data from Salas et al. (2005) and additional 361 cancerous mitochondrial genomes retrieved from the literature were studied as well. To this end, a total of 1,235 cancerous mtDNA sequences were considered in this study (supplementary tables S2 and S3, Supplementary Material online), which were recruited from patients in Asia, Europe, and North America and were composed of 19 types of cancer (astrocytoma, bladder, breast, colorectal, esophageal, glioblastoma, gastric, head and neck, hepatocellular, leukemia, lung, medulloblastoma, nasopharyngeal, oral, ovarian, pancreas, prostate, renal, and thyroid cancer).

Determining Germline Variation in Natural Human Population

To determine the mutational spectra of mtDNA variation in the general populations, 3,696 complete mitochondrial

genomes from the general human populations were retrieved from GenBank (available on June 2009). The mutational spectrum of mtDNA (germline) variation in the general populations was reanalyzed and used as a baseline for comparison. After removing those sequences with poor quality, 3,639 mitochondrial genomes were used in the subsequent analyses. All mtDNA sequences were aligned by Kalign program (Lassmann et al. 2009) according to the rCRS. DNADIST program in PHYLIP 3.69 package (Felsenstein 2009) was used to calculate the distance matrix, followed by the creation of a Neighbor-Joining (NJ) phylogenetic tree using NEIGHBOR with *Pan troglodytes* (NC_001643) as an outgroup. DNAPARS was used to infer the mtDNA sequences at the inner nodes of the NJ phylogenetic tree, and we recorded the variants occurring on the terminal branches as the private ones in the general population (supplementary table S4, Supplementary Material online).

Data Analyses

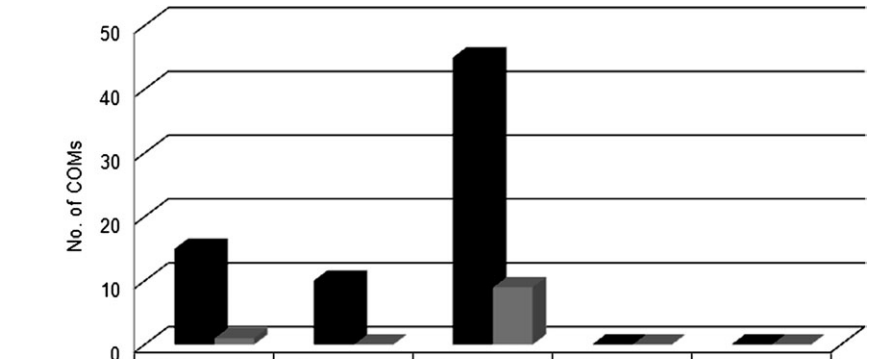
To estimate the potential selective constraints, numbers of synonymous (*S*) and nonsynonymous (*N*) substitutions in the mtDNA protein-coding regions from cancerous tissue samples and the general human populations were counted, respectively. Noticeably, during the analysis of *N*/*S* ratio in each type of cancer, we only considered the cancer type with cancerous mitochondrial genomes outnumbering 30, for a smaller number of cancerous mitochondrial genomes may result in an extremely low incidence of synonymous and/or nonsynonymous mutations and thus easily bias the result. Combination of mutations (COMs, which refers to two or more mutations that recurred in two or more tumors) were calculated according to Zhidkov et al. (2009), and the reported mtDNA data considered in this study were reassessed by using the phylogenetic analysis. This approach could be achieved by extensively comparing the mutation pattern of certain mtDNA sequence under survey with the reconstructed phylogenetic tree of human mtDNAs worldwide, which has been proven to be of great help in distilling potential problems in reported mtDNA data (Bandelt et al. 2001; Yao et al. 2003, 2004, 2006, 2009; Bandelt et al. 2005; Salas et al. 2005; Kong et al. 2008). Specifically, to pinpoint the possibly artificial recombination that was likely introduced by sample contamination or mix-up, the recently proposed phylogenetic method and mutation scoring system were adopted (Kong et al. 2008). To shed more light on the potential effect of data quality on our understanding of mtDNA mutation pattern in cancer studies, ratios of the number of somatic mutations (*Ns*) to the number of the individuals with somatic mutation(s) (*Ni*) were calculated.

Results and Discussion

In the present study, a total of 186 mitochondrial genomes, representing the primary cancerous and adjacent normal tissues of 93 patients with esophageal cancer, were obtained and deposited into GenBank under accession numbers JF824812–JF824997. The new mitochondrial genome data and the cancerous somatic mutations were displayed by way of phylogenetic tree (supplementary fig. S1, Supplementary Material online). As a result, a total of 64 cancerous somatic mutations were detected here (supplementary table S5, Supplementary Material online), among which 18 are nonsynonymous and 11 are synonymous that distribute across the whole genome. Out of the 93 patients, 40 present somatic mutations in their cancerous tissues (40/93), a proportion similar with the previous study on mitochondrial genomes in esophageal cancer (11/20 [Tan et al. 2006]; 12/31 [Gochhait et al. 2008]).

By respectively calculating the ratio of the number of nonsynonymous substitution (*N*) to the number of synonymous substitution (*S*) in the 93 cancerous mitochondrial genomes and those from the general human population, our result revealed that the *N*/*S* ratio in cancerous mitochondrial genomes (1.64) was significantly higher than that in the general populations (0.50; $P = 0.002$). This ratio increased to 2.00 if we considered our previously reported mtDNA data from breast and colorectal cancer (Wang et al. 2007, 2011). The observed significantly higher *N*/*S* value in our esophageal cancerous mtDNAs suggests that the selective pressure on the cancerous mtDNAs is different from that on the mtDNAs from the general human populations. Although the observed high *N*/*S* ratio of cancerous mtDNAs could be explained as the signature of relaxed negative selection, possibility of positive selection could not be ruled out completely. Estimation of neutrality index by the McDonald and Kreitman test (McDonald and Kreitman 1991, by using 24 mitochondrial genomes of chimpanzee as outgroup) revealed that the value of cancerous mtDNAs (2.25) is similar but larger than that of mtDNAs in the matched normal tissues (2.15), indicating the mtDNAs in both tissues being driven by purifying selection, albeit with different degree. Taken together, the mutational pattern observed in the esophageal cancerous mtDNAs might be best explained as the result of the relaxation of negative selection.

Intriguingly, our result is similar to that of Stafford and Chen-Quin (2010) but at odds with the conclusion of Zhidkov et al. (2009). To shed light on the fundamental reasons underlying this discrepancy, we performed a series of analyses to test the possibility of whether the conflict was raised by the problems in data quality. By repeating the analyses performed in Zhidkov et al. (2009) based on the same data set that, however, was corrected by using the phylogenetic approach, we detected only one COMs, significantly fewer than the 15 COMs detected in the data used by Zhidkov et al. (2009; fig. 1 and supplementary table S6, Supplementary Material online). A similar pattern was obtained when the cancerous mitochondrial genome data studied in Salas et al. (2005) were considered: A total of ten COMs were obtained by analyzing the original data set, but no COMs were observed after the reassessed data set were considered (fig. 1 and supplementary table S6, Supplementary Material online). Therefore, it is most likely that most of the previously observed COMs were in fact attributed to problems of data quality, for example, sample mix-up or contamination from the exogenous DNA. Indeed, we found no COMs



	Zhidkov et al.'s data	Salas et al.'s data	All collected data	This study	Data from our laboratory
Original data	15	10	45	0	0
Corrected data	1	0	9	0	0

Fig. 1. Estimation of COMs based on different cancerous mtDNA data sets. "Zhidkov et al.'s data" refers to the data analyzed in Zhidkov et al. (2009); "Salas et al.'s data" means the data considered in Salas et al. (2005); "All collected data" indicates all reported cancerous mtDNA considered in the present study (supplementary tables S2 and S3, Supplementary Material online); "Data from our laboratory" represents the cancerous mitochondrial genome data generated by our laboratory, including a total of 123 cancerous mitochondrial genomes from 93 esophageal cancer patients (this study), 10 breast cancer patients (Wang et al. 2007), and 20 colorectal cancer patients (Wang et al. 2011). "Original data" means the original data retrieved from the literature, whereas "Corrected data" indicates the data corrected by using the phylogenetic approach.

in our newly generated esophageal cancerous mitochondrial genome data (fig. 1). Our reappraising results also indicated that a number of the observed somatic mutations in the mtDNA data set considered in Stafford and Chen-Quin (2010) are likely spurious and attributed to the contamination of exogenous mtDNAs (supplementary table S1, Supplementary Material online). It is then expected that taking this flawed mtDNA data set into consideration would lead to a decrease of the *N/S* value. Indeed, the *N/S* ratio turns out to be 1.32 by using the original data set considered in Stafford and Chen-Quin (2010), much lower than the value (1.64) estimated on the basis of esophageal cancer in this study. Intriguingly, this value is similar to the *N/S* value (1.35) based on all the originally reported cancerous mitochondrial genomes (supplementary table S4, Supplementary Material online).

To obtain a general understanding of the potential effect of poor data quality on the mutational spectrum, we have counted the number of somatic mutations (*Ns*) and the number of the individuals with somatic mutations (*Ni*; see "Data Analyses" in Materials and Methods). By comparing ratios of *Ns*/*Ni* between our own data and those retrieved from the literature, we found that the ratio value of our new data was 1.44, which remained stable (1.47; $P = 1.000$) even after we included our previously reported cancerous mtDNA data from 10 breast cancer patients (Wang et al. 2007) and 20 colorectal cancer patients (Wang et al. 2011). This result, however, is significantly lower than that (2.27) of the reported cancerous mtDNA data from the literature ($P = 0.027$; fig. 2). We corrected the seeming errors in the reported data by using the phylogenetic approach (Bandelt et al. 2005; Salas et al. 2005; Yao et al. 2009), the *Ns*/*Ni* ratio decreases to 1.70, a value similar to our data (1.70 vs. 1.44; $P = 0.45$; fig. 2). Of note is that the corrected mtDNA mutation spectrum (*Ns*/*Ni* ratio) is comparable with our own data (fig. 2), indicating the presence of authentic cancerous somatic mutations in the previous cancer studies albeit some of them most likely suffered from problems in data quality. So far, it is evident that incomplete or biased observation would be easily introduced even when only a small proportion of the cancerous mtDNA data were suffered from potential spurious variation, and this problem

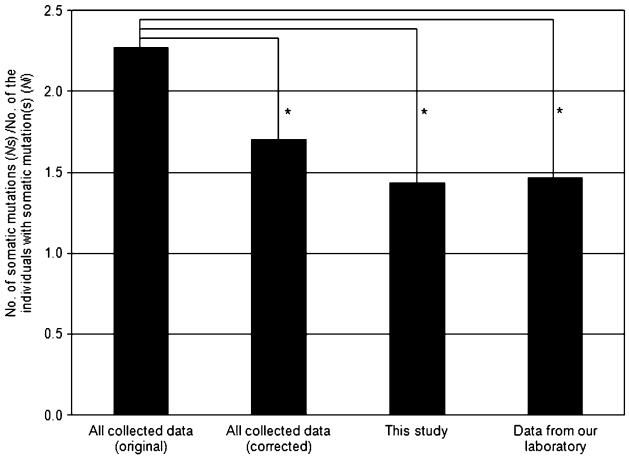


Fig. 2. Comparison of *Ns*/*Ni* ratios between our cancerous mitochondrial genome data and those retrieved from the literature. "All collected data" refers to all the cancerous mtDNA considered in the present study (available in supplementary tables S2 and S3, Supplementary Material online); "original" and "corrected" mean the original data and the data corrected by the phylogenetic approach, respectively; "This study" refers to the 93 esophageal cancerous mitochondrial genome data generated by this study; "Data from our laboratory" means the cancerous mitochondrial genome data generated by our laboratory, including a total of 123 cancerous mitochondrial genomes from 93 esophageal cancer patients (this study), 10 breast cancer patients (Wang et al. 2007), and 20 colorectal cancer patients (Wang et al. 2011). The *P* value of the comparison between B and C is 0.45. $^*P < 0.05$.

Table 1. Distribution of Synonymous and Nonsynonymous mtDNA Mutations in 1,328 mtDNAs from Cancer and 3,639 mtDNAs from the General Population.

	Natural Population			Cancer				
Gene	*N*	*S*	*N/S*	No. of Mutations/bp of Gene	*N*	*S*	*N/S*	*P* Value[a]
MT-ND1	184	309	0.60	0.03	24	5	4.80	1.69×10^{-6}*
MT-ND2	145	372	0.39	0.04	27	19	1.42	3.97×10^{-5}*
MT-ND3	39	105	0.37	0.03	4	6	0.67	0.467
MT-ND4	145	483	0.30	0.03	25	18	1.39	2.23×10^{-6}*
MT-ND4L	14	84	0.17	0.03	9	1	9.00	1.82×10^{-6}*
MT-ND5	352	650	0.54	0.03	34	21	1.62	1.34×10^{-4}*
MT-ND6	91	203	0.45	0.03	9	5	1.80	1.62×10^{-2}*
MT-CYB	294	412	0.71	0.04	29	15	1.93	2.50×10^{-3}*
MT-CO1	116	475	0.24	0.03	22	17	1.29	1.23×10^{-6}*
MT-CO2	82	213	0.39	0.03	9	8	1.13	5.01×10^{-2}
MT-CO3	122	275	0.44	0.05	31	9	3.44	1.01×10^{-8}*
MT-ATP6	272	212	1.28	0.04	23	7	3.29	3.52×10^{-2}*
MT-ATP8	64	87	0.74	0.02	1	4	0.25	0.402
Complex I	970	2206	0.44	0.03	132	75	1.76	7.09×10^{-15}*
Complex III	294	412	0.71	0.04	29	15	1.93	2.50×10^{-3}*
Complex IV	320	963	0.33	0.03	62	34	1.82	1.29×10^{-14}*
Complex V	336	299	1.12	0.04	24	11	2.18	8.21×10^{-2}

NOTE.—*N*, nonsynonymous substitutions; *S*, synonymous substitutions.

[a] *P* value was calculated by Fisher's exact test.

could be largely avoided if the phylogenetic approach was fully appreciated (Yao et al. 2009).

To further determine whether the relaxed selective pressure is a common hallmark among different types of cancer, somatic mutation data of mitochondrial genomes from 1,328 cancerous tissue samples (including our new data) were analyzed and compared with that from the general human populations. Then, by respectively calculating the *N/S* ratio in the cancerous mitochondrial genomes (including the phylogenetically corrected data from the literature as well as our newly generated data) and those from the general human population, our result revealed that the *N/S* ratio of the entire coding region of cancerous mtDNA was 1.83, which was still significantly higher than that (0.50) from the general populations ($P = 3.42 \times 10^{-14}$; table 1). Further analysis revealed that, with the exception of *MT-ND3*, *MT-CO2*, and *MT-ATP8*, the *N/S* ratio of each gene in cancerous mitochondrial genome was significantly higher than that of the general populations (table 1). Intriguingly, a similar significant difference was also observed between the mtDNAs from the cancerous tissues, and the general populations after the genes were allocated into the corresponding respiratory chain complexes: Complex I ($P = 7.09 \times 10^{-15}$), Complex III ($P = 2.50 \times 10^{-3}$), and Complex IV ($P = 1.29 \times 10^{-14}$) (table 1). We classified mitochondrial genomes by cancer type and recalculated their *N/S* ratios independently; our results showed that, for virtually all types of cancer under analysis, their *N/S* ratios were consistently greater than 1 (fig. 3). These similar mutational patterns observed among different types of cancer suggested that cancerous somatic mtDNA mutations were likely accumulated under a common selective pressure, which might be best explained as the relaxation of negative selection. Further evidence in support of this viewpoint came from the observation that almost all the pinpointed cancerous somatic mutations distributed uniformly across the mitochondrial genome (fig. 4).

Since a number of somatic mutations identified in the cancerous tissues were also observed in various normal populations, suggesting them likely to be polymorphic. Excluding these potential polymorphic variants, determined by comparing with the reported variation listed in the

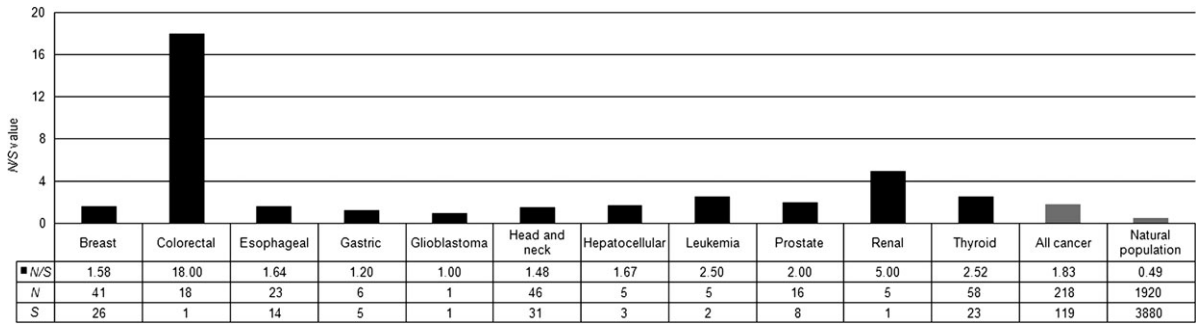


FIG. 3. Assessing the *N/S* values of the cancerous mitochondrial genomes and those from the general populations. Only cancer types with cancerous mitochondrial genomes greater than 30 were considered here. "*N*" indicates nonsynonymous substitution, whereas "*S*" refers to synonymous substitution.

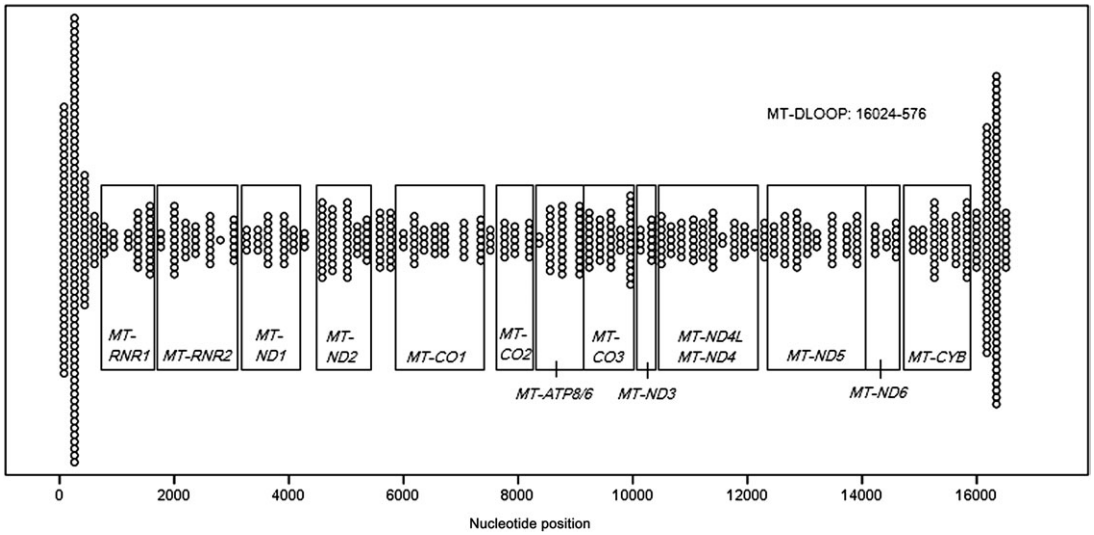


FIG. 4. Distribution of somatic mutations on mitochondrial genome of the cancerous tissue. Each circle represents one mutation in the corresponding region of mtDNA.

mtDB database (http://www.mtdb.igp.uu.se/), from the analyses revealed a similar pattern as observed above: The *N*/*S* ratio of our esophageal cancer mtDNA data increased from 1.64 to 4.33, whereas the total *N*/*S* ratio of all cancerous mtDNA data turned out to be 4.68 (supplementary tables S6 and S7, Supplementary Material online). Of note is that most of the protein-coding genes (except for *MT-ND3*, *MT-ATP6*, and *MT-ATP8*) still showed significantly greater *N*/*S* ratios than those in natural populations, and the previously observed uniform distribution of the cancerous somatic mutations across the mitochondrial genome remained unchanged (supplementary fig. S2, Supplementary Material online).

In conclusion, by analyzing 186 newly obtained mitochondrial genomes and 1,235 reported cancerous mtDNAs, our results suggest that the observed mutational pattern on the mtDNAs of esophageal cancer shall be best explained as the result of the relaxation of negative selection. This pattern is not restricted to esophageal cancer but observed in different types of cancer under consideration, which is in good agreement with the observation that aerobic glycolysis, instead of mitochondrial respiration, plays the key role in generating energy in cancer cells (Vander Heiden et al. 2009). Our result echoes the proposal that the switch in energy metabolism shall be treated as a seventh hallmark of cancer (Hanahan and Weinberg 2011). In retrospect, the observed high frequency of somatic mtDNA mutations in cancer once was treated as increasing evidence in support of the important role of mtDNA in tumorigenesis (cf. Wallace 2005; Brandon et al. 2006; Chatterjee et al. 2006; Hanahan and Weinberg 2011). However, our results challenge this viewpoint by showing that most of cancerous somatic mtDNA mutations were accumulated as a result of the relaxed negative selection, likely due to the impairment of mitochondrial oxidative phosphorylation with the switch of energy metabolism in cancer cells (Warburg 1956; Chandra and Singh 2011). Nonetheless, it is still possible that some specific mtDNA mutations may confer beneficial potential to cancer cells during its development or metastasis (Ishikawa et al. 2008), which plausibly consist of a limited proportion of the distilled somatic mtDNA mutations and could not be detected by the methods employed here. Remarkably, our study also provides solid evidence demonstrating that the problems in some of the published cancerous mtDNA data sets, introduced mainly by sample mix-up or contamination of exogenous mtDNA, well explain the previously conflicting viewpoints of the selective pressure on the cancerous mtDNA, with the spurious somatic mtDNA mutations either leading to the presence of COMs (Zhidkov et al. 2009) or significantly decreasing the *N*/*S* value (Stafford and Chen-Quin 2010), thus offering a paradigm emphasizing the importance of data quality in affecting our understanding on the role of mtDNA in tumorigenesis.

Supplementary Material

Supplementary tables S1–S7 and figures S1 and S2 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

Special thanks to Drs Dong-Dong Wu, Wen-Zhi Wang, Hai-Bing Xie, and Prof. Peng Shi for their helpful suggestions. This work was supported by grants from the Chinese Academy of Sciences (KSCXZ-EW-J-23), Science and Technology Bureau of Kunming (08H130301), National Natural Science Foundation of China (30900797 and 30925021), and Yunnan Province (2009CI119).

References

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet.* 23:147.

Bandelt HJ, Achilli A, Kong QP, Salas A, Lutz-Bonengel S, Sun C, Zhang YP, Torroni A, Yao YG. 2005. Low "penetrance" of

phylogenetic knowledge in mitochondrial disease studies. *Biochem Biophys Res Commun.* 333:122–130.
Bandelt HJ, Lahermo P, Richards M, Macaulay V. 2001. Detecting errors in mtDNA data by phylogenetic analysis. *Int J Legal Med.* 115:64–69.
Bandelt HJ, Salas A, Taylor RW, Yao YG. 2009. Exaggerated status of "novel" and "pathogenic" mtDNA sequence variants due to inadequate database searches. *Hum Mutat.* 30:191–196.
Brandon M, Baldi P, Wallace DC. 2006. Mitochondrial mutations in cancer. *Oncogene* 25:4647–4662.
Chandra D, Singh KK. 2011. Genetic insights into OXPHOS defect and its role in cancer. *Biochim Biophys Acta.* 1807:620–625.
Chatterjee A, Mambo E, Sidransky D. 2006. Mitochondrial DNA mutations in human cancer. *Oncogene* 25:4663–4674.
Coller HA, Khrapko K, Bodyak ND, Nekhaeva E, Herrero-Jimenez P, Thilly WG. 2001. High frequency of homoplasmic mitochondrial DNA mutations in human tumors can be explained without selection. *Nat Genet.* 28:147–150.
Crespi BJ, Summers K. 2006. Positive selection in the evolution of cancer. *Biol Rev Camb Philos Soc.* 81:407–424.
Felsenstein J. 2009. PHYLIP (phylogeny inference package) version 3.69. Seattle (WA): Department of Genome Sciences, University of Washington.
Fendt L, Zimmermann B, Daniaux M, Parson W. 2009. Sequencing strategy for the whole mitochondrial genome resulting in high quality sequences. *BMC Genomics* 10:139.
Gochhait S, Bhatt A, Sharma S, Singh YP, Gupta P, Bamezai RN. 2008. Concomitant presence of mutations in mitochondrial genome and p53 in cancer development—a study in north Indian sporadic breast and esophageal cancer patients. *Int J Cancer.* 123:2580–2586.
Hanahan D, Weinberg RA. 2011. Hallmarks of cancer: the next generation. *Cell* 144:646–674.
Ishikawa K, Takenaga K, Akimoto M, Koshikawa N, Yamaguchi A, Imanishi H, Nakada K, Honma Y, Hayashi J. 2008. ROS-generating mitochondrial DNA mutations can regulate tumor cell metastasis. *Science* 320:661–664.
Kong QP, Bandelt HJ, Sun C, et al. (12 co-authors). 2006. Updating the East Asian mtDNA phylogeny: a prerequisite for the identification of pathogenic mutations. *Hum Mol Genet.* 15:2076–2086.
Kong QP, Salas A, Sun C, Fuku N, Tanaka M, Zhong L, Wang CY, Yao YG, Bandelt HJ. 2008. Distilling artificial recombinants from large sets of complete mtDNA genomes. *PLoS One* 3:e3016.
Kong QP, Yao YG, Sun C, Bandelt HJ, Zhu CL, Zhang YP. 2003. Phylogeny of east Asian mitochondrial DNA lineages inferred from complete sequences. *Am J Hum Genet.* 73:671–676.
Lassmann T, Frings O, Sonnhammer EL. 2009. Kalign2: high-performance multiple alignment of protein and nucleotide sequences allowing external features. *Nucleic Acids Res.* 37:858–865.
McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in Drosophila. *Nature* 351:652–654.
Nowell PC. 1976. The clonal evolution of tumor cell populations. *Science* 194:23–28.
Palanichamy MG, Sun C, Agrawal S, Bandelt HJ, Kong QP, Khan F, Wang CY, Chaudhuri TK, Palla V, Zhang YP. 2004. Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. *Am J Hum Genet.* 75:966–978.
Salas A, Yao YG, Macaulay V, Vega A, Carracedo A, Bandelt HJ. 2005. A critical reassessment of the role of mitochondria in tumorigenesis. *PLoS Med.* 2:e296.
Stafford P, Chen-Quin EB. 2010. The pattern of natural selection in somatic cancer mutations of human mtDNA. *J Hum Genet.* 55:605–612.
Sun C, Kong QP, Palanichamy MG, Agrawal S, Bandelt HJ, Yao YG, Khan F, Zhu CL, Chaudhuri TK, Zhang YP. 2006. The dazzling array of basal branches in the mtDNA macrohaplogroup M from India as inferred from complete genomes. *Mol Biol Evol.* 23:683–690.
Tan DJ, Chang J, Liu LL, Bai RK, Wang YF, Yeh KT, Wong LJ. 2006. Significance of somatic mutations and content alteration of mitochondrial DNA in esophageal cancer. *BMC Cancer* 6:93.
van Oven M, Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat.* 30:E386–E394.
Vander Heiden MG, Cantley LC, Thompson CB. 2009. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* 324:1029–1033.
Wallace DC. 2005. Mitochondria and cancer: Warburg addressed. *Cold Spring Harb Symp Quant Biol.* 70:363–374.
Wang CY, Li H, Hao XD, Liu J, Wang JX, Wang WZ, Kong QP, Zhang YP. 2011. Uncovering the profile of somatic mtDNA mutations in Chinese colorectal cancer patients. *PLoS One* 6:e21613.
Wang CY, Wang HW, Yao YG, Kong QP, Zhang YP. 2007. Somatic mutations of mitochondrial genome in early stage breast cancer. *Int J Cancer.* 121:1253–1256.
Warburg O. 1956. On respiratory impairment in cancer cells. *Science* 124:269–270.
Yao YG, Bravi CM, Bandelt HJ. 2004. A call for mtDNA data quality control in forensic science. *Forensic Sci Int.* 141:1–6.
Yao YG, Kong QP, Salas A, Bandelt HJ. 2008. Pseudomitochondrial genome haunts disease studies. *J Med Genet.* 45:769–772.
Yao YG, Macauley V, Kivisild T, Zhang YP, Bandelt HJ. 2003. To trust or not to trust an idiosyncratic mitochondrial data set. *Am J Hum Genet.* 72:1341–1346.
Yao YG, Salas A, Bravi CM, Bandelt HJ. 2006. A reappraisal of complete mtDNA variation in East Asian families with hearing impairment. *Hum Genet.* 119:505–515.
Yao YG, Salas A, Logan I, Bandelt HJ. 2009. mtDNA data mining in GenBank needs surveying. *Am J Hum Genet.* 85:929–933.
Zhidkov I, Livneh EA, Rubin E, Mishmar D. 2009. MtDNA mutation pattern in tumors and human evolution are shaped by similar selective constraints. *Genome Res.* 19:576–580.
Zhou S, Kachhap S, Sun W, et al. (18 co-authors). 2007. Frequency and phenotypic implications of mitochondrial DNA mutations in human squamous cell cancers of the head and neck. *Proc Natl Acad Sci U S A.* 104:7540–7545.