Phylogenetic and Evolutionary Analysis of Chinese Leishmania Isolates Based on Multilocus Sequence Typing

Chun-Ying Zhang, Xiao-Jun Lu, Xiao-Qing Du, Jun Jian, Ling Shu, Ying Ma*

Introduction

Leishmaniasis is a debilitating infectious disease that has a variety of clinical forms. In China, visceral leishmaniasis (VL) is the most common symptom, and L. donovani and/or L. infantum are the likely pathogens. In this study, multilocus sequence typing (MLST) of five enzyme-coding genes (fh, gpdh, icd, mpi, pgd) and two conserved genes (hsp70, lack) was used to investigate the phylogenetic relationships of Chinese Leishmania strains. Concatenated alignment of the nucleotide sequences of the seven genes was analyzed and phylogenetic trees were constructed using neighbor-joining and maximum parsimony models. A set of additional sequences from 25 strains (24 strains belong to the L. donovani complex and one strain belongs to L. gerbilli) were retrieved from GenBank to infer the molecular evolutionary history of Leishmania from China and other endemic areas worldwide. Phylogenetic analyses consolidated Chinese Leishmania into four groups: (i) one clade A population comprised 13 isolates from different foci in China, which were pathogenic to humans and canines. This population was subdivided into two subclades, clade A1 and clade A2, which comprised sister organisms to the remaining members of the worldwide L. donovani complex; (ii) a population in clade B consisted of one reference strain of L. turalica and five Chinese strains from Xinjiang; (iii) clade C (SELF-7 and EJNI-154) formed a population that was closely related to clade B, and both isolates were identified as L. gerbilli; and (iv) the final group, clade D, included Sauroleishmania (LIZRD and KKG-E) and was distinct from the other strains. We hypothesize that the phylogeny of Chinese Leishmania is associated with the geographical origins rather than with the clinical forms (VL or CL) of leishmaniasis. To conclude, this study provides further molecular information on Chinese Leishmania isolates and the Chinese isolates appear to have a more complex evolutionary history than previously thought.

Abstract

Leishmaniasis is a debilitating infectious disease that has a variety of clinical forms. In China, visceral leishmaniasis (VL) is the most common symptom, and L. donovani and/or L. infantum are the likely pathogens. In this study, multilocus sequence typing (MLST) of five enzyme-coding genes (fh, gpdh, icd, mpi, pgd) and two conserved genes (hsp70, lack) was used to investigate the phylogenetic relationships of Chinese Leishmania strains. Concatenated alignment of the nucleotide sequences of the seven genes was analyzed and phylogenetic trees were constructed using neighbor-joining and maximum parsimony models. A set of additional sequences from 25 strains (24 strains belong to the L. donovani complex and one strain belongs to L. gerbilli) were retrieved from GenBank to infer the molecular evolutionary history of Leishmania from China and other endemic areas worldwide. Phylogenetic analyses consolidated Chinese Leishmania into four groups: (i) one clade A population comprised 13 isolates from different foci in China, which were pathogenic to humans and canines. This population was subdivided into two subclades, clade A1 and clade A2, which comprised sister organisms to the remaining members of the worldwide L. donovani complex; (ii) a population in clade B consisted of one reference strain of L. turalica and five Chinese strains from Xinjiang; (iii) clade C (SELF-7 and EJNI-154) formed a population that was closely related to clade B, and both isolates were identified as L. gerbilli; and (iv) the final group, clade D, included Sauroleishmania (LIZRD and KKG-E) and was distinct from the other strains. We hypothesize that the phylogeny of Chinese Leishmania is associated with the geographical origins rather than with the clinical forms (VL or CL) of leishmaniasis. To conclude, this study provides further molecular information on Chinese Leishmania isolates and the Chinese isolates appear to have a more complex evolutionary history than previously thought.

Citation: Zhang C-Y, Lu X-J, Du X-Q, Jian J, Shu L, et al. (2013) Phylogenetic and Evolutionary Analysis of Chinese Leishmania Isolates Based on Multilocus Sequence Typing. PLOS ONE 8(4): e63124. doi:10.1371/journal.pone.0063124

Editor: Dan Zilberstein, Technion - Israel Institute of Technology, Israel

Received November 30, 2012; Accepted March 29, 2013; Published April 30, 2013

Copyright: © 2013 Zhang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by grants from the National Nature Science Foundations of China(81071425). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: majiying72@hotmail.com

Infant VL infections dramatically increased in Jiashi County, Xinjiang [11].

The Leishmania species in China are complex. Various clinical forms exist, ranging from cutaneous to fatal VL, and they are caused by human/canine pathogenic Leishmania species. Some species infect gerbils and lizards, but these are nonpathogenic to humans [12]. According to different epidemiological characteristics, leishmaniasis in China has been classified into three types [13]: plain, hill and desert foci. The different hosts and vectors for the parasites in distinct foci make unveiling the relationship between Leishmania strains difficult. For example, L. donovani is generally considered to be the pathogen of plain type leishmaniasis, and only sporadic cases of this infection have been reported since the 1980s. In Gansu and Sichuan provinces, hill type cases of leishmaniasis are prevalent, and most patients are children aged less than 10 years. In this instance, domestic dogs are the key reservoir. Xinjiang is a mixed area with two epidemiological types, plain and desert [9,11]. The plain type is endemic in the oases of the plains of Kashgar City, where children aged less than 5 years are susceptible. The desert type is distributed in the desert regions of Xinjiang. Two species were found in this area, one is L. donovani...
complex that causes autochthonous kala-azar in infants, and the other species is *L. tropicana*, which is nonpathogenic to humans.

Research groups have investigated the phylogenetic relationship of the Chinese *Leishmania* strains using internal transcribed spacer 1 (ITS1) sequencing [14] and kinetoplast cytochrome oxidase II (COII) gene sequencing [15]. Some studies have classified Chinese isolates as *L. infantum* [16,17], while other research groups suggest that Chinese Leishmania isolates do not form a monophyletic group (such as *L. donovani* or *L. infantum*) due to the presence of a novel, undescribed species of *Leishmania* [14,15,18]. However, most of these studies included single species isolates or isolates from a limited number of geographic regions. Therefore, the taxonomic diversity and phylogenetic relationship of Chinese *Leishmania* strains remain unclear.

As morphological distinction of *Leishmania* species is impractical, the immunological and DNA-based criteria are used to identify the different strains. The universally accepted standard procedure for characterizing and identifying strains of *Leishmania* is multilocus enzyme electrophoresis (MLEE) [19]. However, the drawbacks of MLEE have restricted its usage [20,21] and this method has been challenged by many alternative methods for species discrimination using molecular markers. These include restriction fragment length polymorphism (RFLP) analysis of the specific gene or gene fragment, such as *cpb/gpd3* [22], multilocus microsatellite typing (MLMT) fragment analysis [16,17,23], and multilocus sequencing typing (MLST) which targets conserved genes [24,25].

The introduction of molecular approaches, particularly the numerous typing methodologies with various targets, has called for a revision of *Leishmania* taxonomy [26,27]. Multilocus sequence typing (MLST) have been used to provide new insights into the population genetics, taxonomy and evolutionary history of *Leishmania*, and multiple specific MLST targets are available for *Leishmania* species [24,25,28].

Based on the published targets for the sub-genus of *Leishmania*, we combined five enzyme-coding genes (*fh*, *g6pdh*, *icd*, *mpi* and *pgd*) and two conserved genes (*hsp70* and *lack*) to differentiate the Chinese *Leishmania* isolates and to investigate their phylogenetic relationships.

**Materials and Methods**

**Ethics Statement**

Details of the WHO codes, sources, geographical origins and clinical manifestations of the 28 *Leishmania* isolates used in this study are listed in Table 1. Fig. 1 shows the different locations from where the Chinese strains had been isolated. We included 22 *Leishmania* strains isolated from different foci in China and 6 WHO *Leishmania* reference strains. Two strains, MCAN/CN/11/1101 and MCAN/CN/11/1102, were isolated from two dogs in China in 2011. The other 26 strains included in this study were kindly provided by Professor Wang Junyun at the Shanghai Municipal Center for Disease Control and Prevention, and has been previously analyzed by Wang Y et al. [29]. For the isolation of MCAN/CN/11/1101 and MCAN/CN/11/1102, oral informed consent was obtained from the guardians of the two dogs. The original animal work that produced the samples was conducted in accordance with guidelines of the Chinese Council on Animal Care, as approved by the Animal Care Committee of West China Hospital, Sichuan University.

**Parasite Culture and Genomic DNA Preparation**

The parasites were isolated as promastigotes on Novy-MacNeal-Nicolle biphasic culture medium and grown in Medium 199 supplemented with 15–20% heat-inactivated fetal bovine serum at 22–24°C. The promastigotes were harvested at room temperature by centrifugation at 4000 rpm for 10 min and washed three times with NET buffer (50 mM Tris-HCl, 125 mM EDTA, 50 mM NaCl). Total genomic DNA was extracted immediately using a proteinase K-phenol/chloroform method [30]. DNA was suspended in TE buffer (10 mM Tris-HCl, 1 mM EDTA) and stored at –20°C until required.

**PCR Amplification and Gene Sequencing**

Five enzyme-coding genes [24,25] (fumarate hydratase (*fh*), glucose-6-phosphate dehydrogenase (*g6pdh*), isocitrate dehydrogenase (*icd*), mannose phosphate isomerase (*mpi*), and 6-phosphogluconate dehydrogenase (*pgd*)), and two conserved genes [31,32] (heat shock protein 70 (*hsp70*) and *Leishmania* homolog of receptors for activated protein kinase C (*lack*)) were used in this study. The primers were listed in Table 2. After an initial denaturation step for 10 min at 95°C, DNA samples were processed through 30 cycles of 1 min at 95°C, 1 min at the annealing temperature (Tₐ) indicated in Table 2, 90 s at 72°C, followed by a terminal elongation step of 10 min at 72°C. Amplification products were examined by electrophoresis on 1.5% agarose gels and stained with ethidium bromide.

PCR products of the expected size were sequenced directly in both strands by ABI 3730 DNA Sequencers (Applied Biosystems) using the same primer pairs in the initial PCR reactions. The sequences were assembled with the aid of ContigExpress (http://www.contigexpress.com/), and a consensus sequence was generated for each strain. Apart from strain WDD23 for *g6pdh*, no gaps or insertions was detected in the sequences obtained from the strains used in this study. The sequence gene were deposited in GenBank (accession numbers shown in Table 1).

**Nucleotide Sequence Analysis**

The sequences were aligned using ClustalW program in the MEGA 5.0 package [33]. The aligned matrix from this procedure was manually rechecked and minor adjustments were made if required. The transition/transversion ratios were calculated using the Kimura-2-Parameter and the synonymous and nonsynonymous substitution rates of selection were calculated using the modified Nei-Gojobori method (Jukes-Cantor distance model) with the MEGA 5.0 package.

**Protein Sequence Analysis**

Molecular weight, overall charge, and charge variation according to theoretical isoelectric point (pI) for each partial protein sequence were determined using online resources (http://web.expasy.org/compute_pi/).

**Phylogenetic Analysis**

Phylogenetic trees were constructed with each single gene sequence of *fh*, *g6pdh*, *icd*, *mpi*, *pgd*, *hsp70* and *lack* with one or no heterozygous genes. The DNA sequences of strains with one or more heterozygous sites were deconstructed into two (or several) individual sequences. As the similarity in sequences produced poor resolution result of a single gene tree, the concatenated alignment of the nucleotide sequences for the five enzyme-coding genes or all seven genes were analyzed. Prior to phylogenetic analyses, the most appropriate model of evolution, GTR+I+F+G, was determined using ModelTest v.0.1.1 [34] under the Akaike Information Criterion (AIC). Different statistical methods (for example, the maximum likelihood tree, UPGMA tree and neighbor-joining tree) were calculated in MEGA 5.0 [33]. The support of monophyletic groups was assessed by the bootstrap method with
1,000 replicates. Simultaneously, the PAUP (version 4.0b) [35] program was used to construct phylogenetic trees by maximum parsimony analyses and to perform bootstrap with 1,000 replications.

A set of additional sequences of 25 strains (shown in Table 1, 24 strains belong to the *L. donovani* complex and one strain belong to *L. gerbilli* ) were retrieved from GenBank for the global analyses. The retrieved sequences were from the studies of Mauricio *et al.* [24] and Zemanova *et al.* [25]. Alignment simulations were performed using the ClustalW program. Phylogenetic and molecular evolutionary analyses were conducted using MEGA 5.0 and PAUP 4.0b as mentioned above.

### Results

#### PCR Amplification of the Seven Genes

Using the primers listed in Table 2, a single product of the expected size for each gene was amplified from each of the Chinese Leishmania strains and the WHO reference strains. Three strains (LIZRD, KXG-E and M2903) only yielded the gene sequences of *hsp70* and *lack*. As only the region between the primers was sequenced, only partial section of the entire coding region of each gene was analyzed. The lengths of the sequences, the nucleotide positions of the start and the end of the sequences analyzed for the genes, their G+C contents and the transition/transversion ratios (Ts/Tv) were given in Table 2. The nonsynonymous/synonymous substitution ratios (dN/dS) varied, though they were all below 1.00 for seven genes. On a Z-test for selection, all the genes were neutral.

#### DNA Sequences

Two major groups of aligned sequences were noted. The first group included 13 Chinese isolates and a WHO *L. donovani* reference strain DD8. The second group included five Chinese isolates (strain KXG-Y, KXG-R, KXG-57, KXG-11 and QITAI-15) and a WHO *L. turanica* reference strain 3720 as reported in a previous study [29]. When the two groups were compared, several nucleotide differences at numerous sites were found, so the DNA sequences were analyzed for the two groups, respectively.

Table 3 shows the polymorphisms within the 13 Chinese isolates. The number of single nucleotide polymorphic (SNP) sites varied among the genes. For example, there were seven mutations in the 1671 bp (7/1671) section of the *fh* gene, of which four were silent and three resulted in altered amino acid residues. Six SNPs (6/1578) were observed in *g6pdh*, and no alternative amino acid was identified. Two mutations were detected in both the *icd* and *pgd* genes, no amino acid change was observed in *icd*, and one position was involved with an amino acid change in *pgd*. Five diversity nucleotides were displayed in *mpi*, with one change in an amino acid residue. A low level of polymorphism was revealed for *hsp70* among the isolates. There were three heterozygous sites for strain 801. A high degree of similarity was observed among the *lack* gene sequences, and no amino acid changes were detected.

For the six strains (3720, KXG-Y, KXG-R, KXG-57, KXG-11 and QITAI-15) of *L. turanica* mentioned above, the sequences had a high degree of similarity, apart from the synonymous mutations that occurred at a few mutant sites.
| Name          | Species         | Location     | Host     | Disease | Accession numbers in GenBank |
|---------------|-----------------|--------------|----------|---------|------------------------------|
| 9044          | Strains         | Shandong, China | Human   | VL      | JX021310, JX021329           |
| SC3           | Strains         | Nanping, Sichuan | Human   | VL      | JX021311, JX021330           |
| SC3           | Strains         | Nanping, Sichuan | Canine  | JX021312, JX021331           |
| SC9           | Strains         | Nanping, Sichuan | Canine  | JX021313, JX021332           |
| fh            | Strains         | Karamay, Xinjiang | Human   | CL      | JX021314, JX021333           |
| icd           | Strains         | Karamay, Xinjiang | Human   | CL      | JX021315, JX021334           |
| mpi           | Strains         | Karamay, Xinjiang | Canine  | JX021316, JX021335           |
| pgd           | Strains         | Karamay, Xinjiang | Canine  | JX021317, JX021336           |
| Hsp70         | Strains         | Wenchang, China | Canine  | JX021318, JX021337           |
| lack          | Strains         | Wenchang, China | Canine  | JX021319, JX021338           |
| fh            | Strains         | Kashgar, Xinjiang | Human   | VL      | JX021320, JX021340           |
| icd           | Strains         | Kashgar, Xinjiang | Human   | VL      | JX021321, JX021341           |
| mpi           | Strains         | Kashgar, Xinjiang | Canine  | JX021322, JX021342           |
| pgd           | Strains         | Kashgar, Xinjiang | Canine  | JX021323, JX021343           |
| Hsp70         | Strains         | Kashgar, Xinjiang | Canine  | JX021324, JX021344           |
| lack          | Strains         | Kashgar, Xinjiang | Canine  | JX021325, JX021345           |
| fh            | Strains         | Wenxian, Gansu | Human   | VL      | JX021326, JX021346           |
| icd           | Strains         | Wenxian, Gansu | Human   | VL      | JX021327, JX021347           |
| mpi           | Strains         | Wenxian, Gansu | Canine  | JX021328, JX021348           |
| pgd           | Strains         | Wenxian, Gansu | Canine  | JX021329, JX021349           |
| Hsp70         | Strains         | Wenxian, Gansu | Canine  | JX021330, JX021350           |
| lack          | Strains         | Wenxian, Gansu | Canine  | JX021331, JX021351           |
| fh            | Strains         | Karamay, Xinjiang | Rodent  | JX021332, JX021352           |
| icd           | Strains         | Karamay, Xinjiang | Rodent  | JX021333, JX021353           |
| mpi           | Strains         | Karamay, Xinjiang | Rodent  | JX021334, JX021354           |
| pgd           | Strains         | Karamay, Xinjiang | Rodent  | JX021335, JX021355           |
| Hsp70         | Strains         | Karamay, Xinjiang | Rodent  | JX021336, JX021356           |
| lack          | Strains         | Karamay, Xinjiang | Rodent  | JX021337, JX021357           |
| fh            | Strains         | Turpan, Xinjiang | Lizard  | JX021338, JX021358           |
| icd           | Strains         | Turpan, Xinjiang | Lizard  | JX021339, JX021359           |
| mpi           | Strains         | Turpan, Xinjiang | Lizard  | JX021340, JX021360           |
| pgd           | Strains         | Turpan, Xinjiang | Lizard  | JX021341, JX021361           |
| Hsp70         | Strains         | Turpan, Xinjiang | Lizard  | JX021342, JX021362           |
| lack          | Strains         | Turpan, Xinjiang | Lizard  | JX021343, JX021363           |
| fh            | Strains         | Ejin, Inner Mongolia | Rodent  | JX021344, JX021364           |
| icd           | Strains         | Ejin, Inner Mongolia | Rodent  | JX021345, JX021365           |
| mpi           | Strains         | Ejin, Inner Mongolia | Rodent  | JX021346, JX021366           |
| pgd           | Strains         | Ejin, Inner Mongolia | Rodent  | JX021347, JX021367           |
| Hsp70         | Strains         | Ejin, Inner Mongolia | Rodent  | JX021348, JX021368           |
| lack          | Strains         | Ejin, Inner Mongolia | Rodent  | JX021349, JX021369           |
| fh            | Strains         | Blitar, Indonesia | Rodent  | JX021350, JX021370           |
| icd           | Strains         | Blitar, Indonesia | Rodent  | JX021351, JX021371           |
| mpi           | Strains         | Blitar, Indonesia | Rodent  | JX021352, JX021372           |
| pgd           | Strains         | Blitar, Indonesia | Rodent  | JX021353, JX021373           |
| Hsp70         | Strains         | Blitar, Indonesia | Rodent  | JX021354, JX021374           |
| lack          | Strains         | Blitar, Indonesia | Rodent  | JX021355, JX021375           |
| fh            | Strains         | Kashgar, Xinjiang | Rodent  | JX021356, JX021376           |
| icd           | Strains         | Kashgar, Xinjiang | Rodent  | JX021357, JX021377           |
| mpi           | Strains         | Kashgar, Xinjiang | Rodent  | JX021358, JX021378           |
| pgd           | Strains         | Kashgar, Xinjiang | Rodent  | JX021359, JX021379           |
| Hsp70         | Strains         | Kashgar, Xinjiang | Rodent  | JX021360, JX021380           |
| lack          | Strains         | Kashgar, Xinjiang | Rodent  | JX021361, JX021381           |
| fh            | Strains         | Wenxian, Gansu | Rodent  | JX021362, JX021382           |
| icd           | Strains         | Wenxian, Gansu | Rodent  | JX021363, JX021383           |
| mpi           | Strains         | Wenxian, Gansu | Rodent  | JX021364, JX021384           |
| pgd           | Strains         | Wenxian, Gansu | Rodent  | JX021365, JX021385           |
| Hsp70         | Strains         | Wenxian, Gansu | Rodent  | JX021366, JX021386           |
| lack          | Strains         | Wenxian, Gansu | Rodent  | JX021367, JX021387           |
| fh            | Strains         | Ejin, Inner Mongolia | Lizard  | JX021368, JX021388           |
| icd           | Strains         | Ejin, Inner Mongolia | Lizard  | JX021369, JX021389           |
| mpi           | Strains         | Ejin, Inner Mongolia | Lizard  | JX021370, JX021390           |
| pgd           | Strains         | Ejin, Inner Mongolia | Lizard  | JX021371, JX021391           |
| Hsp70         | Strains         | Ejin, Inner Mongolia | Lizard  | JX021372, JX021392           |
| lack          | Strains         | Ejin, Inner Mongolia | Lizard  | JX021373, JX021393           |
| fh            | Strains         | Blitar, Indonesia | Lizard  | JX021374, JX021394           |
| icd           | Strains         | Blitar, Indonesia | Lizard  | JX021375, JX021395           |
| mpi           | Strains         | Blitar, Indonesia | Lizard  | JX021376, JX021396           |
| pgd           | Strains         | Blitar, Indonesia | Lizard  | JX021377, JX021397           |
| Hsp70         | Strains         | Blitar, Indonesia | Lizard  | JX021378, JX021398           |
| lack          | Strains         | Blitar, Indonesia | Lizard  | JX021379, JX021399           |
### Table 1. Cont.

| Name   | Strain      | Species        | Location          | Host   | Disease | Accession numbers in GenBank |
|--------|-------------|----------------|-------------------|--------|---------|------------------------------|
|        |             |                |                   |        |         | fh | G6pdh | icd | mpi | pgd | Hsp70 | lack |
| L100   | MHOM/ET/72/L100 | L. aethiopica | Wollo, Ethiopia    | Human  | CL      | KC763812 |
| K-27   | MHOM/SU/74/K27  | L. tropica     | Azerbaijan         | Human  | CL      | KC763809 |
| S5ASKH | MHOM/SU/73/S5ASKH | L. major      | Turkmenistan       | Human  | CL      | KC763810 |
| M2903  | MHOM/IR/75/M2903 | L. braziliensis | Brazil            | Human  | CL      | KC763811 |

(B) The strains whose sequences were taken from GenBank.

*The species identification is disputed dependent on the technique applied.*

*Two strains LIZRD and KXG-E did not amplify the five enzyme-coding genes (fh, G6pdh, icd, mpi and pgd), so there is no accession numbers in GenBank.*

*For the 6 WHO reference strains, we did not submit the corresponding sequences obtained here to GenBank because the sequences have been studied and submitted to GenBank by other researchers. Meanwhile, we aligned the sequences of the enzyme-coding genes and hsp70 gene for the 6 reference strains we determined with those sequences retrieved from GenBank, the sequences for the genes we obtained were the same with those in GenBank, respectively.*

*Strains whose sequences taken from GenBank were studied using MLST by Mauricio et al., 2006, and Zemanova et al., 2007, there are no sequences of hsp70 and lack genes from the corresponding strains.*

*doi:10.1371/journal.pone.0063124.t001*
Table 2. Primers used for gene sequencing and PCR.

| Target | Primer name | Primer sequence | Y (Tm) °C | Expected length (bp) | Sequences for alignment (begin to end) | Length of the sequences (bp) | G+C contents | Ts/Tv b | Reference |
|--------|-------------|-----------------|-----------|----------------------|----------------------------------------|----------------------------|---------------|---------|-----------|
| fh     | fh-f        | AGCGTCTTGTTTCCTCCA | 60        | 1707                 | 7–1677                                 | 1671                       | 63%           | 3.65    | Zemanova et al., 2007 |
| fh     | fh-r        | GAGCCCGTTAGAGGAGG |           |                      |                                        |                           |               |         |           |
| g6pdh  | g6pdh-f     | ATGTCCGAAGGAGAGCT | 50        | 1689                 | 64–1641                                | 1578                       | 55%           | 3.68    | Zemanova et al., 2007 |
| g6pdh  | g6pdh-r     | TCACACGCTTATTGGAGGAA |            |                      |                                        |                           |               |         |           |
| iad    | iad-f       | ATGTTCGCGCATTGTTCCG | 55        | 1308                 | 58–1260                                | 1203                       | 59.5%         | 3.73    | Zemanova et al., 2007 |
| iad    | iad-r       | TTAGCCGCTACGCCCTTT |           |                      |                                        |                           |               |         |           |
| mpi    | mpi-f       | ATGTCGAGCTGTAAGGCT | 55        | 1266                 | 58–1212                                | 1155                       | 61%           | 2.38    | Zemanova et al., 2007 |
| mpi    | mpi-r       | CTACCTGCTGCTCAGTCC |           |                      |                                        |                           |               |         |           |
| pgd    | pgd-f       | GAAAGAATCCCTTTATCTCYATG | 60     | 1440                 | 40–1431                                | 1392                       | 61.5%         | 2.41    | Mauricio et al., 2006 |
| pgd    | pgd-r       | GGAACGGGTGGCCGGGCGC |           |                      |                                        |                           |               |         |           |
| hsp70  | hsp70-f     | GACGGTGCGCTGCTACCTCAC | 60     | 1380                 | 40–1317                                | 1278                       | 65%           | 1.36    | Fraga et al., 2010    |
| hsp70  | hsp70-r     | CGGAGGAGGCTGCTCTGGTACATC |         |                      |                                        |                           |               |         |           |
| lack   | Lack-f      | ACCATGACACTACGAGGTCACCCT | 41     | 942                  | 49–897                                 | 849                        | 66.2%         | 2.18    | Gonzalez-Aseguinolaza et al., 1999 |
| lack   | Lack-r      | TACTCGCGCCTCGAGAT |           |                      |                                        |                           |               |         |           |

af is forward and r is reverse.

Ts/Tv, transition/transversion ratios.

doi:10.1371/journal.pone.0063124.t002
The initial tree constructed based on individual gene comparisons did not produce discrepancy tree topologies (data not shown). The tree shown in Fig.2 was constructed based on the concatenated sequences of \textit{hsp70} and \textit{lack} using the neighbor-joining method in MEGA5.0 package. The phylogenetic analysis of \textit{hsp70} and \textit{lack} showed that 13 Chinese \textit{Leishmania} isolates grouped as a population in clade A, which were well separated from other strains. Clade A was further divided into two subpopulations, clade A1 and clade A2. Five Chinese strains formed a group (clade B) with WHO \textit{L. turanica} reference strain 3720, as reported previously [29]. Two distinct groups were also produced, one including SELF-7 and EJNI-154 (clade C) which

| Gene | Strain | Nucleotide position/Nucleotide | Allele | pI | MMass |
|------|--------|--------------------------------|--------|----|-------|
| \textit{hsp70} | DD8, SC6 | C C C C G A T | 1 | 6.34 | 61609.10 |
| | 771, 9044, XKG-LIU, XKG-XU | C T T G G A C | 2 | 6.34 | 61609.10 |
| | KXG-65 | T T T G G A T | 3 | 6.94 | 52725.10 |
| | SC9 | C C C C A A T | 4 | 6.31 | 61486.88 |
| | 1101, 1102, 801, WDD23 | C C C C A A T | 4 | 6.34 | 61636.12 |
| | K56, 8801 | C C C C A T T | 5 | 6.44 | 61620.17 |
| | | 303 423 522 1413 1580 1595 1596 | | | |
| \textit{gpdh} | DD8 | C T T T T G C | 1 | 5.73 | 59771.14 |
| | 771, 9044, XKG-LIU, XKG-XU, XKG-65 | C C T C T G C | 2 | 5.73 | 59771.14 |
| | SC6 | C C C C T G C | 3 | 5.73 | 59771.14 |
| | SC9, 1101, 1102, 801, 8801 | T T T C T G C | 4 | 5.73 | 59771.14 |
| | WDD23 | T T T C C G C | 5 | 5.65 | 59577.84 |
| | K56 | T T T C T C C | 6 | 5.73 | 59785.17 |
| | | 315 894 | | | |
| \textit{icd} | DD8, 771, 9044, SC6, XKG-LIU, XKG-XU, XKG-65 | C C | 1 | 7.72 | 45087.56 |
| | SC9, 1101, 1102, 801, WDD23, K56, 8801 | T A | 2 | 7.72 | 45087.56 |
| | | 258 499 600 756 1059 | | | |
| \textit{mpi} | DD8, 771, 9044, XKG-LIU, XKG-XU | T G G T G | 1 | 5.91 | 42727.90 |
| | KXG-65 | T G G T A | 2 | 5.91 | 42727.90 |
| | SC6 | T G A T G | 3 | 5.91 | 42727.90 |
| | SC9, 1101, 1102, 801, WDD23 | C A G C G | 4 | 6.18 | 42726.96 |
| | 801, K56, 8801 | C G G C G | 5 | 5.91 | 42727.90 |
| | | 840 952 | | | |
| \textit{pgd} | DD8 | C A | 1 | 6.24 | 50858.08 |
| | 9044, XKG-LIU, XKG-XU, SC6 | C G | 2 | 6.01 | 50859.06 |
| | SC9, 1101, 1102 | C G | 2 | 6.01 | 50859.06 |
| | 801, K56, 8801, WDD23 | C G | 2 | 6.01 | 50859.06 |
| | 771, XKG-65 | A G | 3 | 6.01 | 50859.06 |
| | | 385 887 1010 1161 1193 | | | |
| \textit{hsp70} | DD8, 9044, SC6, XKG-LIU, XKG-XU | C C C C C | 1 | 5.28 | 46951.92 |
| | 771 | C C T C C C | 1/2 | 5.23 | 46814.78 |
| | KXG-65 | C T C C C C | 2 | 5.23 | 46814.78 |
| | SC9, 1101, 1102, 8801, WDD23, K56 | T C C C C C | 3 | 5.28 | 46951.92 |
| | 801 | T C C T C/A | 3/4 | 5.65 | 46806.92 |
| | | 600 759 | | | |
| \textit{lack} | DD8 | C T | 1 | 6.04 | 31206.98 |
| | 771, 9044, SC6, | C T | 2 | 6.04 | 31206.98 |
| | SC6, XKG-LIU, XKG-XU | C C | 3 | 6.04 | 31206.98 |
| | SC9, 1101, 1102, 801, WDD23, K56, 8801 | C T | 1 | 6.04 | 31206.98 |

*Nucleotide differences in sequences of individual genes for 14 strains in clade A are shown, along with the position of the base mutations. In bold: sites that cause an amino acid polymorphism. See deposited sequences for additional details.

doi:10.1371/journal.pone.0063124.t003

Phylogenetic Trees

The initial tree constructed based on individual gene comparisons did not produce discrepancy tree topologies (data not shown). The tree shown in Fig.2 was constructed based on the concatenated sequences of \textit{hsp70} and \textit{lack} using the neighbor-joining method in MEGA5.0 package.
were closely related to clade B in the tree, and the other population (clade D), consisting of LIZRD and KXG-E, was separated from the other Chinese isolates.

Phylogenetic tree was also constructed based on the concatenated sequences of the seven genes (Supplement Fig.1). Only 25 isolates were included in the analysis: three isolates, LIZRD, KXG-E and M2903, were excluded as only hsp70 and lack genes were amplified. The tree topologies calculated by using different statistical methods or substitution models did not differ, so only the tree constructed by the neighbor-joining method based on Kimura 2-Parameter distance matrices in the MEGA 5.0 package (Supplement Fig.1) was presented, bootstraps with 1,000 replicates were applied. Three major clades (clades A, B and C) were also recognized in the neighbor-joining tree based on the combined sequences. Clade A included a WHO L. donovani reference strain DD8 and Chinese isolates which were pathogenic to humans and canines, and the Clade A can be subdivided into two general subclades: subclade A1 included the majority of strains from humans and sandflies; subclade A2 contained the strains isolated from humans and canines. The coexistence of different disease phenotypes (VL and CL) in the same species complex (clade A) was noted in this study, more strikingly, the same genotype from different disease phenotypes. In a previous study, incongruence between the Leishmania genotypes and the VL/CL disease phenotype of the isolates had also been noted [7]. Clade B included five Chinese Leishmania strains and a WHO reference L. turanica strain 3720. A distinct group, clade C, was detected for strain SELF-7 and EJNI-154, which were identified as L. gerbilli.

We also investigated the phylogenetic relationships between Chinese Leishmania isolates in conjunction with isolates from other geographical regions. The trees generated by the entire dataset are displayed in Fig. 3 (and Supplement Fig.2). In both the neighbor-joining and maximum parsimony trees, 13 Chinese isolates formed two subclades which were separate from other strains of the L. donovani complex isolated from Europe, India and Africa. The branches that clustered in the tree reflected the geographical origins of the Leishmania strains. The six isolates from Xinjiang, Sichuan and Shandong Province in China comprised the subclade A1, the other seven isolates from Xinjiang, Sichuan and Gansu provinces in China supported the subclade A2. Additionally, the strains in clade A1 were close to the Indian and Africa isolates, while strains in clade A2 were close to the European isolates.

Discussion

Multilocus sequence typing (MLST) has the potential to provide a substantial contribution to the understanding of the epidemiol-
ogy, transmission and phylogenetics of infectious diseases, and the data can then be shared by depositing or withdrawing information from the GenBank, EMBL or DDBJ databases [36]. In this study, seven gene targets (fh, g6pdh, icd, mpi, pgd, hsp70 and lack) were described for *Leishmania* strains in China, and the phylogenetic relationships and evolution of Chinese *Leishmania* isolates were investigated.

Despite the low level of SNP variation observed in our study, the existing differences of concatenated sequences reveal significant genetic diversity. The neighbor-joining or maximum parsimony methods identified several clusters when gene sequences of Chinese *Leishmania* strains were aligned with those of the WHO reference strains. In fact, these clusters were also observed in other tree topologies using different models (data not shown), which indicates that the derived groups were robust and not dependent on the choice of evolutionary methods underlying the various tree-building algorithms. Within the *Leishmania* subgenus, there was correlation between intraspecies divergence and the geographic distribution of the species. Within clade A, isolates from Shandong and Karamay of Xinjiang (clade A1) apparently differed from the majority of the closely related strains isolated from Gansu and Kashgar of Xinjiang (clade A2). This further demonstrated that the geographic origin of a strain was a more important predictor of genetic relationship than the type of disease presentation [37].

It has been proposed that environmental changes of human origin may be attributed to population structure and phylogenetic diversity [38]. The hosts play important roles in determining the outcome of leishmaniasis. The parasites in different hosts may develop different escape responses that result in variable mutation rates. Minor differences exist between strain SC6 from a patient and SC9 from a canine may have been a result of the host/reservoir origins. More isolates from humans and canines in Sichuan Province are needed to illustrate the phenomenon.

---

**Figure 3. Phylogenetic trees constructed based on the sequences of the fh, g6pdh, icd, mpi and pgd genes for 25 isolates in this study and 24 isolates of the *Leishmania donovani* complex from other studies.** Neighbor-joining tree was constructed from five enzyme genes for the 49 isolates using MEGA 5.0 software. The Kimura-2-parameter method was used. Numbers above branches correspond to bootstrap values based on 1,000 replicates. The strains were designated by their names (See Table 1 for more details).

doi:10.1371/journal.pone.0063124.g003
In clade B, five isolates from Xinjiang, an autonomous region of China, were highlighted by substantial genetic differentiations from L. donovani complex at the species level. The result of seven conserved genes indicated that the five strains were closely related to strain 3720, which was classified as L. turanica. Our results support the conclusion of previous studies [29] that these five strains were grouped as L. turanica. Two strains (SELF-7 and EJN1-154) in clade C have been previously identified as L. gerbilli [29]. The sequences of gpdh, ica, and mpi for strain SELF-7 and EJN1-154 were aligned with the WHO L. gerbilli reference strain, MRHO/CN/1960/Gerbilli, in the analysis, and strain SELF-7, EJN1-154 were corroborated as L. gerbilli. Additionally, five enzyme genes did not amplify from three isolates (LIZRD, KXG-E and M2903). Strain LIZRD and KXG-E were grouped into clade D in Fig. 2 when the hsp70 and lack genes were analyzed. The strain KXG-E was genetically similar to Sauroleishmania strain LIZRD described by Wang et al. [29]. As Sauroleishmania and Leishmania are genetically significantly different [39], the changes in the primer annealing sites for strains LIZRD and KXG-E may have caused the failure in amplification. New primers located in the intragenic region of the corresponding enzyme-coding genes were designed and only part of the sequence for the single enzyme-coding gene was amplified. For strain M2903 of L. braziliensis, a New World species that belongs to a different sub-genus (Viannia), it was found to be distantly related to the Old World species. The enzyme-coding genes of this strain were not amplified with the primers used in this study as a result of the inappropriate primers.

Based on the trees constructed from our database and other sequences of the L. donovani complex worldwide, the results demonstrate that Chinese Leishmania isolates (clade A1 and A2) are sister to other members of the L. donovani complex. L. donovani complex species from China were segregated into two groups: clade A1 was most closely related to L. donovani strains from India and Africa, whereas clade A2 species were found to be related to the European strains of L. infantum. In the tree shown in Fig. 3, the clade A1 strains were placed between the Leishmania species from India and Africa. Previous finding has shown that Indian and African L. donovani strains differ considerably, and our results were in accordance with the previous study [40]. Thus, we further provide an evidence of a robust grouping of Chinese Leishmania when compared with the worldwide isolates, which provides a strong indication that the isolates from China have a more complex evolutionary history than previously thought. This study provides a basis for a further understanding of the phylogeny and evolutionary history of Leishmania parasitic species in China.

Whether Chinese Leishmania isolates are actually grouped as an undescribed species on the basis of new insights is a matter of debate. In the present study, sampling bias may exist for phylogenies, even though we tried to overcome this problem by selecting strains from diverse geographic origins and hosts. In addition, sequence variants may exist among different cells in a given population when uncloned cultures are used for DNA isolation. The direct DNA isolation of Leishmania from clinical samples would avoid this potential complication in the analysis. Further studies based on the current dataset of Chinese Leishmania isolates should be undertaken to gain further insight into the classification and evolution of Leishmania parasites. A systematic comparison of a statistically significant number of worldwide isolates would be an ideal way to elucidate the evolutionary profile of Leishmania species.

Supporting Information

Figure S1 Phylogenetic trees constructed based on sequences of the fh, gpdh, ica, mpi, pgd, hsp70 and lack genes for 25 Leishmania isolates in this study. The neighbor-joining unrooted tree was constructed using MEGA 5.0 software. The Kimura-2-parameter method was used. Numbers above branches correspond to bootstrap values based on 1,000 replicates. The strains were designated by their names (see Table 1 for more details). (TIF)

Figure S2 Phylogenetic trees constructed based on the sequences of the fh, gpdh, ica, mpi and pgd genes for the 25 isolates in this study and 24 isolates of the Leishmania donovani complex from other studies. Maximum parsimony tree constructed with the sequences of five enzyme genes for 49 isolates using the PAUP 4.0b program. The trees were rooted with L. tropica (K27) and L. major (5ASKH). Numbers above branches correspond to bootstrap values based on 1,000 replicates. The strains were designated by their names (see Table 1 for more details). (TIF)

Author Contributions

Conceived and designed the experiments: CZ, XL, YM. Performed the experiments: CZ, XL, LS, YM. Analyzed the data: CZ, XL, XD, JJ, YM. Contributed reagents/materials/analysis tools: XL, YM. Wrote the paper: CZ, YM.

References

1. Ashford RW (2000) The leishmaniases as emerging and reemerging zoonoses. Int J Parasitol 30: 1269–1281.
2. Alvar J, Vezel ID, Bern C, Herrera M, Desjeux P, et al. (2012) Leishmaniasis worldwide and global estimates of its incidence. PLoS One 7: e35871.
3. Cruz I, Nieto J, Moreno J, Canavate C, Desjeux P, et al. (2006) Leishmania/HIV co-infections in the second decade. Indian J Med Res 125: 337–338.
4. Guan LR, Qu JQ, Chai JJ (2000) Leishmania in China: present status, prevalence and some suggestions in its control. Endem Dis Bull 15: 49–52. (In Chinese with English abstract)
5. Guan LR, Zuo XP, Yinamu. (2005) Reemergence of visceral leishmaniasis in Kashgar Prefecture, Xinjiang. Chin J Parasitol Parasitic Dis 21: 285. (In Chinese with English abstract)
6. Zhou XX, Le S, Yang GJ, Kristensen TK, Bergquist NR, et al. (2009) Spatial epidemiology in zoonotic parasitic diseases: insights gained at the 1st International Symposium on Geospatial Health in Lijiang, China, 2007. Parasit Vectors 2: 10.
7. Waki K, Dutta S, Ray D, Kolli BK, Akman L, et al. (2007) Transmembrane molecules for phylogenetic analyses of pathogenic protozoa: Leishmania-specific informative sites in hydrophilic loops of trans-endoplasmic reticulum N-acetylglucosamine-1-phosphate transferase. Eukaryot Cell 6: 198–210.
8. Lu DM, Hu XS, Qiao ZD (2001) Analysis of Leishmania species and strains from China by RAPD technique. Chin J Parasitol Parasit Dis 19: 290–293. (In Chinese with English abstract)
9. Guan LR (2009) Present situation of visceral leishmaniasis and prospect for its control in China. Chin J Parasitol Parasitic Dis 27: 394–397. (In Chinese with English abstract)
10. Yisilayin O, Hou YY (2011) Retrospective analysis of prevalence of visceral leishmaniasis in Xinjiang from 2003 to 2010. Bull Dis Control Prev 26: 3–6. (In Chinese with English abstract)
11. Wang JY, Gao CH, Yang YT, Chen HT, Zha XH, et al. (2010) An outbreak of the desert sub-type of zoonotic visceral leishmaniasis in Jiashi, Xinjiang Uygur Autonomous Region, People’s Republic of China. Parasitol Int 59: 331–337.
12. Guan LR, Jin CF, Xu YX (1982) Studies on Leishmania in big gerbils and some suggestions in its control. Endem Dis Bull 15: 49–52. (In Chinese with English abstract)
13. Guan LR, Wang J, Hu YD (1976) The epidemiological types of China Kala-azar and theri significance in disease control. Study Epidemiic Diseases control 3: 225–232. (In Chinese with English abstract)
14. Yang BB, Gao XG, Hu XS, Zhang JG, Liao L, et al. (2010) Species discrimination and phylogenetic inference of 17 Chinese Leishmania isolates based on internal transcribed spacer 1 (ITS1) sequences. Parasitol Res 107: 1049–1065.
