Livestock trade history, geography, and parasite strains: the mitochondrial genetic structure of Echinococcus granulosus in Argentina.

Permalink
https://escholarship.org/uc/item/5692d7g7

Journal
The Journal of parasitology, 90(2)

ISSN
0022-3395

Authors
Haag, KL
Ayala, FJ
Kamenetzky, L
et al.

Publication Date
2004-04-01

DOI
10.1645/ge-173r

License
https://creativecommons.org/licenses/by/4.0/ 4.0

Peer reviewed
LIVESTOCK TRADE HISTORY, GEOGRAPHY, AND PARASITE STRAINS: THE MITOCHONDRIAL GENETIC STRUCTURE OF ECHINOCOCCUS GRANULOSUS IN ARGENTINA

K. L. Haag, F. J. Ayala*, L. Kamenetzky*, A. M. Gutierrez†, and M. Rosenzvit†
Departamento de Genética, Universidade Federal do Rio Grande do Sul, Caixa Postal 15053, CEP 91501-970 Porto Alegre, RS, Brazil. e-mail: karen.haag@ufrgs.br

ABSTRACT: A sample of 114 isolates of Echinococcus granulosus (Cestoda: Taeniidae) collected from different host species and sites in Argentina has been sequenced for 391 bp from the mitochondrial cytochrome c oxidase subunit I gene to analyze genetic variability and population structure. Nine different haplotypes were identified, 5 of which correspond to already characterized strains. Analysis of molecular variance and nested clade analysis of the distribution of haplotypes among localities within 3 main geographic regions indicate that geographic differentiation accounts for the overall pattern of genetic variability in E. granulosus populations. Significant geographic differentiation is also present when the sheep strain alone is considered. Our results suggest that geographic patterns are not due to actual restricted gene flow between regions but are rather a consequence of past history, probably related to the time and origin of livestock introduction in Argentina.

Echinococcus (Cestoda: Taeniidae) includes a group of 4 species, i.e., Echinococcus vogeli, E. oligarthrus, E. multilocularis, and E. granulosus. They require 2 hosts to complete their cycles, i.e., a carnivore, in which the adult occurs, and a herbivore, in which the metacestode develops. Hermaphroditic adult parasites are believed to reproduce mainly by self-fertilization (Thompson, 1995), although there is some evidence of cross-fertilization (Lymbery et al., 1997; Haag et al., 1998). The virtual absence of outcrossing, associated with the great asexual proliferation potential of the metacestode, may lead to a high degree of strain differentiation, which is well known in at least 1 of the species, i.e., E. granulosus (Thompson and McManus, 2002).

Echinococcus granulosus strains have different intermediate host specificities. Whereas adults consistently infect canids, metacestodes from different strains appear to be adapted to distinct species of domestic and wild herbivorous hosts, including sheep, cattle, pig, horses, and wild cervids. However, the host specificities of a given strain vary enormously. The sheep strain is the least specific, using a wide range of intermediate host species, including several domestic ungulates, humans, and macropods, although fertile cysts (metacestodes undergoing active asexual reproduction) are more frequent in sheep. It is also the most geographically widespread E. granulosus strain, occurring in all continents. Some Echinococcus strains are highly divergent, as has been shown by genetic, morphologic, and developmental characters (see Eckert and Thompson, 1997 for a review), leading to the notion that they should be regarded as different species (Bowles et al., 1995; Thompson et al., 1995; Thompson and McManus, 2002).

Cystic hydatid disease (the infection caused by the metacestode of E. granulosus) is considered a reemerging zoonosis in several countries, such as Bulgaria, Kazakhstan, and the Republic of China (Eckert et al., 2000). Although South America has several endemic areas (Cabrera et al., 1995; Schantz et al., 1995; Eckert et al., 2000), little is known about parasite genetic variation in these regions. In Argentina, where most of the present data have been collected, hydatid disease is endemic in several provinces, affecting not only livestock such as cattle (7% infected), sheep (12.5%), pigs (9.8%), and goats (6.0%) (data provided by Servicio Nacional de Sanidad y Calidad Agropecuaria [SENASA], 1987–1996) but also wild animals such as hares (Schantz and Lord, 1972). In some regions of Neuquen, Chubut, and Rio Negro Provinces, human prevalences of 6,000–14,900 per 100,000 inhabitants have been reported (Larrieu et al., 1999). Kamenetzky et al. (2002) have shown the presence of a great genetic variability among 147 E. granulosus isolates collected from different host species distributed throughout the south of South America, but mainly within Argentina. Our interpretation suggests the presence of at least 5 strains plus some minor genetic variants. In this article, we analyzed a 391-bp-long fragment of the mitochondrial cytochrome c oxidase subunit I (CO1) from the same Argentine isolates within a population genetics framework. Our goals were to evaluate the genetic variability and population structure of E. granulosus in the region and to ascertain some of their possible causes.

MATERIALS AND METHODS

Sampling and sequencing

One hundred and fourteen E. granulosus parasites (101 metacestodes and 13 adults) were isolated from different hosts and provinces in Argentina. For some purposes, isolates were grouped into 3 geographic areas (south, central, and north; see Fig. 1), according to the information received from field workers during the collection of parasites from their hosts. The distribution of isolates among hosts and geographic areas is shown in Table I. After DNA extraction, a 391-bp segment of the mitochondrial cytochrome c oxidase subunit I (CO1) from the same Argentine isolates was amplified by polymerase chain reaction (PCR), purified, and sequenced automatically, as described previously (Kamenetzky et al., 2002).

In this article, the term “strain” always refers to a genetically homogeneous group of isolates belonging to the same species, i.e., E. granulosus; “haplotype” designates the mitochondrial sequence used to discriminate strains, and “variants” are haplotypes with minor genetic differences within a strain.

Statistical analyses

Sequences were aligned using the Pileup program from the GCG Package (Genetics Computer Group, Version 9.1). Polymorphism estimates were calculated with Arlequin Version 2.0 (Schneider et al., 2000). The total mitochondrial CO1 genetic variability in the whole sample of 114 isolates and also in some subsamples (see Results) was

Received 20 June 2003; revised 31 July 2003; accepted 13 August 2003.
* Department of Ecology and Evolutionary Biology, University of California Irvine, 321 Steinhaus Hall, Irvine, California 92697.
tested for significant genetic differentiation among geographic regions. For this, we used the analysis of molecular variance (AMOVA) statistics (Excoffier et al., 1992) developed in Arlequin Version 2.0 (Schneider et al., 2000). A second approach was followed to decompose the total genetic variability, using nested clade analysis (reviewed in Templeton, 1998). Networks were estimated by a statistical parsimony procedure using the TCS Version 1.13 software (Clement et al., 2001), nested by the rules described in Crandall (1996), and tested for geographical associations with the program GeoDis Version 2.0 (Posada et al., 2000), using the geographic coordinates of the 36 localities shown in Figure 1.

RESULTS

Five of the 9 CO1 sequences found in the whole sample are identical to already known haplotypes (Bowles et al., 1992), corresponding to 5 distinctive E. granulosus strains. The other 4 sequences included minor genetic differences and, therefore, were referred to as variants of the 5 already known haplotypes (see Fig. 2 and Table II for details). Our data show that whereas the southern and central populations have a predominance of the sheep strain G1 haplotype, the northern population has a higher frequency of haplotypes G1c (sheep strain variant) and G2 (Tasmanian sheep strain). Moreover, no pig strain isolate was found in the north.

Polymorphism estimates were calculated not only from the whole sample but also from 2 other subsamples including: (1) all isolates collected from human hosts or (2) all haplotypes related to the sheep strain (G1, G1a, G1b, G1c, and G2). The results are shown in Table III. The whole sample comprises a five, but divergent, haplotypes, representing the strains (see also Table II). When the major haplotype differences due to strain differentiation are excluded from the analysis, i.e., considering only the sheep strain subsample, nucleotide diversities reduce by 1–2 orders of magnitude and haplotype diversities are reduced to about half. However, it is interesting to note that the northern population appears to be less affected when only the sheep strain subsample is considered. Human hosts from all over Argentina harbor a great variety of parasites. The polymorphism estimates of the human subsample (n = 63) are comparable with those obtained for the entire sample (n = 114).

Indeed, humans are potential hosts for most E. granulosus strains found in this study, with the exception of the pig strain. As a first approximation to the mitochondrial genetic structure of E. granulosus in Argentina, we ran AMOVA using the same sample and subsamples of isolates described previously. As shown in Table IV, geographic differentiation in Argentina is high, but some within-region substructuring occurs (most variation is found within populations), which might represent the presence of several divergent strains. However, we believe that the highly significant Fst values are not merely related to the use of different host species in each geographic region because

![Figure 1](image1.png)

**Figure 1.** Geographic distribution of the 36 Echinococcus granulosus sampling sites (indicated by stars). Their geographic coordinates were used in the nested clade analysis, but for other purposes, the 36 sites were grouped into 3 “populations”, delimited by their latitude: South > 37°S 00’ > Central > 28°S 50’ > North.

![Figure 2](image2.png)

**Figure 2.** Multiple alignment of the 47 polymorphic sites among the 9 CO1 sequences obtained in this study. Haplotypes corresponding to the already known strains are represented by G1–G7 as in Bowles et al. (1992), and their minor variants are indicated by letters (G1a–G1c, and G7a). Vertically oriented numbers indicate the site position.

### Table I. Number of Echinococcus granulosus isolates collected from each host species according to geographic region. Numbers in parentheses are percentages relative to the number of isolates in each geographic region (inner cells) or in the whole sample (right and bottom margins).

<table>
<thead>
<tr>
<th>Hosts</th>
<th>South</th>
<th>Central</th>
<th>North</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>40 (67.8)</td>
<td>5 (21.7)</td>
<td>18 (56.2)</td>
<td>63 (55.3)</td>
</tr>
<tr>
<td>Dog</td>
<td>8 (13.5)</td>
<td>1 (4.35)</td>
<td>4 (12.5)</td>
<td>13 (11.4)</td>
</tr>
<tr>
<td>Cattle</td>
<td>—</td>
<td>8 (34.8)</td>
<td>4 (12.5)</td>
<td>12 (10.5)</td>
</tr>
<tr>
<td>Sheep</td>
<td>4 (6.8)</td>
<td>1 (4.35)</td>
<td>6 (18.8)</td>
<td>11 (9.7)</td>
</tr>
<tr>
<td>Goat</td>
<td>1 (1.7)</td>
<td>2 (8.37)</td>
<td>—</td>
<td>3 (2.6)</td>
</tr>
<tr>
<td>Pig</td>
<td>6 (10.2)</td>
<td>6 (26.1)</td>
<td>—</td>
<td>12 (10.5)</td>
</tr>
<tr>
<td>Total</td>
<td>59 (51.7)</td>
<td>23 (20.2)</td>
<td>32 (28.1)</td>
<td>114</td>
</tr>
</tbody>
</table>
TABLE II. Number of Echinococcus granulosus isolates from each CO1 haplotype per geographic region (S, South; C, Central; N, North). Haplotype frequencies are shown in parentheses. The last 2 columns indicate the host species in which haplotypes were found (Hu, human; Do, dog; Ca, cattle; Sh, sheep; Pi, pig; Go, goat) and their corresponding strains (as reported in Bowles et al., [1992]).

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>GenBank</th>
<th>S (%)</th>
<th>C (%)</th>
<th>N (%)</th>
<th>Total</th>
<th>Hosts</th>
<th>Strain</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>M84661</td>
<td>27 (45.76)</td>
<td>11 (47.83)</td>
<td>2 (6.25)</td>
<td>40 (35.09)</td>
<td>Hu, Do, Ca, Sh, Pi</td>
<td></td>
</tr>
<tr>
<td>G1a</td>
<td>AF455871</td>
<td>5 (8.47)</td>
<td>—</td>
<td>2 (6.25)</td>
<td>7 (6.14)</td>
<td>Hu, Do, Ca, Pi</td>
<td></td>
</tr>
<tr>
<td>G1b</td>
<td>AF458872</td>
<td>2 (3.39)</td>
<td>—</td>
<td>1 (4.35)</td>
<td>3 (2.63)</td>
<td>Hu, Ca</td>
<td></td>
</tr>
<tr>
<td>G1c</td>
<td>AF458873</td>
<td>—</td>
<td>—</td>
<td>11 (34.38)</td>
<td>11 (34.38)</td>
<td>Hu, Ca, Sh</td>
<td></td>
</tr>
<tr>
<td>G2</td>
<td>M84665</td>
<td>—</td>
<td>2 (8.70)</td>
<td>3 (9.38)</td>
<td>5 (4.39)</td>
<td>Hu, Do, Ca</td>
<td></td>
</tr>
<tr>
<td>G6</td>
<td>M84666</td>
<td>21 (35.39)</td>
<td>3 (13.04)</td>
<td>5 (15.62)</td>
<td>29 (25.44)</td>
<td>Hu, Do, Go</td>
<td></td>
</tr>
<tr>
<td>G7</td>
<td>M84667</td>
<td>4 (6.78)</td>
<td>2 (8.70)</td>
<td>—</td>
<td>6 (5.26)</td>
<td>Pi</td>
<td></td>
</tr>
<tr>
<td>G7a</td>
<td>AF758876</td>
<td>—</td>
<td>4 (17.39)</td>
<td>—</td>
<td>4 (3.51)</td>
<td>Do, Pi</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>59</td>
<td>23</td>
<td>32</td>
<td>114</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE III. CO1 polymorphism in the southern, central, and northern Argentine populations, considering the whole Echinococcus granulosus sample, as well as the human and sheep strain subsamples. n, sample size; k, number of alleles; π nucleotide diversity; h haplotype diversity. Standard deviations are shown in parentheses.

<table>
<thead>
<tr>
<th></th>
<th>South</th>
<th>Central</th>
<th>North</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole sample</td>
<td>n 59</td>
<td>23</td>
<td>32</td>
<td>114</td>
</tr>
<tr>
<td></td>
<td>k 5</td>
<td>6</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>π 0.0454 (0.0227)</td>
<td>0.0526 (0.0288)</td>
<td>0.0360 (0.0184)</td>
<td>0.0450 (0.0223)</td>
</tr>
<tr>
<td></td>
<td>h 0.6622 (0.0383)</td>
<td>0.7391 (0.0787)</td>
<td>0.8024 (0.0422)</td>
<td>0.7932 (0.0240)</td>
</tr>
<tr>
<td>Human subsample</td>
<td>n 40</td>
<td>5</td>
<td>18</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>k 4</td>
<td>2</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>π 0.0455 (0.0229)</td>
<td>0.0358 (0.0227)</td>
<td>0.0396 (0.0207)</td>
<td>0.0431 (0.0215)</td>
</tr>
<tr>
<td></td>
<td>h 0.6244 (0.0425)</td>
<td>0.4000 (0.2373)</td>
<td>0.8366 (0.0572)</td>
<td>0.7363 (0.0341)</td>
</tr>
<tr>
<td>Sheep strain subsample*</td>
<td>n 34</td>
<td>12</td>
<td>24</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>k 3</td>
<td>2</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>π 0.0009 (0.0010)</td>
<td>0.0004 (0.0003)</td>
<td>0.0058 (0.0037)</td>
<td>0.0033 (0.0023)</td>
</tr>
<tr>
<td></td>
<td>h 0.3557 (0.0942)</td>
<td>0.1667 (0.1343)</td>
<td>0.6920 (0.0646)</td>
<td>0.6294 (0.0033)</td>
</tr>
</tbody>
</table>

* G1, G1a, G1b, G1c, and G2.
not statistically significant, as tested by a randomization procedure based on 1,000 resamples.

DISCUSSION

The intraspecific genetic variability of flatworms has been little explored. Most genetic studies have focused on strain differences because of their implications in parasite diagnostics and disease control. Moreover, most of our knowledge of E. granulosus genetic variability comes from Australian populations, with a long history of control programs (Thompson and McManus, 2002). One of our previous studies in E. granulosus populations of Argentina (Rosenzvit et al., 1999) showed, by mitochondrial DNA sequences and ribosomal DNA PCR–restriction fragment length polymorphism, the presence of 4 distinct strains in human and domesticated intermediate hosts, namely the sheep, Tasmanian sheep, camel, and pig strains. Recently, a larger dataset has allowed us to confirm the previous results (Kamenetzky et al., 2002) and to find additional variation by sequencing a part of the mitochondrial CO1 gene. In this article we report the analysis of population genetic structure based on these data.

Some of our findings are not new. Thus, the existence of more than 1 E. granulosus strain in a single geographic area has already been reported, i.e., in northwestern China, for example, where the sheep and camel strains were identified (Zhang et al., 1998), and in southern Brazil, where the simultaneous occurrence of sheep and cattle strains was shown as well (Haag et al., 1998). Although our isolates were grouped somewhat arbitrarily into geographic regions, which do not represent real populations in the genetic sense, such a high number of different haplotypes and a marked geographic differentiation have never been reported for E. granulosus. Lymbery et al. (1997) suggest that a possible cause for the lack of genetic structure in E. granulosus populations from Australia is the high mobility of its hosts. The superposition of a sylvatic cycle (maintained by dingos and macropods) and a domestic cycle (with dogs and mainly sheep as hosts) could also contribute to the genetic homogeneity of populations by spreading parasite genes throughout the continent.

In other macroparasites for which population genetic data have been obtained, both the absence and the presence of geographic structure have been explained by the extent of gene flow. Restricted gene flow in genetically structured populations is inferred by the hosts’ sedentary habits, such as soil-dwelling insects (Blouin et al., 1999) and snails (Dybdahl and Lively, 1996; Sire et al., 2001), or with limited dispersal capacities, such as deer compared with domestic ruminants (Blouin et al., 1995). The absence of correlation between geographic and genetic distances for the CO1 gene in populations of hookworms from China was attributed to uneven human host movements and variable effective sizes (Hawdon et al., 2001). Even though historical and demographic hypotheses have been suggested to explain population genetic patterns, no statistic tool has been used to discriminate between them.

Three causes might explain the geographic pattern of E. granulosus genetic differentiation in Argentina: (1) regional differences in animal livestock, (2) actual restricted gene flow, and (3) historical causes, i.e., variations in time of domestic animal (along with their parasites) introductions and regional peculiarities in animal trade or husbandry practices during the last few centuries. The first cause can never be ruled out because regional differences indeed exist. For example, most of our pig strain isolates are found in the central region (see Table II), where pigs are traditionally raised in Argentina. But this does not account for the significant geographic differentiation found within the sheep strain, unless the sheep strain variants are adapted to different livestock species or strains. We did not
consider natural selection on CO1 as a plausible explanation because the nucleotide differences present in the sheep strain haplotype variants do not lead to amino acid substitutions, except for G1b, in which an isoleucine is replaced by a valine. Because the mitochondrial genome is inherited as a single unit, these substitutions could nevertheless be favored by hitchhiking and, for simplicity, we define this possibility as a historical cause, together with random processes such as bottleneck events, or human-directed causes such as parasite introduction with domestic animals.

The geographic distribution of CO1 haplotypes in this study shows that the derived haplotypes G1c and G2 are restricted to the northern population, whereas the ancestral G1 haplotype is more frequent in the central and southern regions. These geographic associations could be best explained by historical causes, as indicated by the discrepancy between Dc and Dn of these tip clades in the nested clade analysis (see Table V). Moreover, the significantly large Dc and small Dn for the average interior minus average tip clades (I − T) within category 2-1 strongly supports the idea that the derived clades occur far from their ancestor. If these regional differences are not a mere sampling artifact, they imply long-distance, historical movements. Most of our northern samples come from an isolated mountainous region called Valle del Tafí and Cumbres Calchaquies, where livestock was introduced a long time ago. These domestic animals probably remained isolated for decades because of their low economic value. Animal husbandry in this region is used exclusively for subsistence.

It would be difficult to date exactly when that isolation occurred and which host species or strains (together with their parasite variants) have been introduced, but there is a fair possibility that the *E. granulosus* G2 haplotype came with sheep imported from Australia at the beginning of the 20th century. The G1c variant could have arisen by mutation from the common sheep strain haplotype (G1), which might have been introduced with European cattle much earlier (the earliest cattle trades occurred in the 16th century), and increased in frequency by genetic drift or selection acting on other mitochondrial genes. The frequency of genetic variants can increase rapidly in *Echinococcus* spp. populations because of their asexual amplification in the metacestode stage, followed by egg dispersion via domestic or wild dog movements. Apparently, because neither dogs nor ruminants from this isolated region disperse to the south, the range of these parasite variants remains restricted.

The explanation for the restricted G7a (pig strain variant) haplotype distribution is less clear. It may be mainly related to the high intermediate host specificity of this strain and to some geographic peculiarities of pig husbandry in Argentina. An analysis of a larger set of pig strain isolates would be necessary to confirm the geographic association and to fully understand its causes. We are now working on the development of nuclear microsatellite markers, which might be useful to analyze population structure and transmission within single strains. Studying the pattern and the causes of the apparently complex genetic structure of *E. granulosus* will permit a better understanding of the disease process and diagnosis as well as enable control measures adapted to each parasite strain.

**ACKNOWLEDGMENTS**

Thanks to PADCT, FAPERGS, CNPq, CAPES, CBAB, CABBIO, and CONICET for supporting our work. We are also indebted to Aldo Mellender de Araújo, John Hawdon, and Don McManus for the review of this manuscript.

**LITERATURE CITED**


**Table V. Geographic clade distances (Dc) and nested clade distances (Dn) of the tip clades showing statistically significant values in the geographic distance analysis (G1c, G2, and G7a). I − T corresponds to the average interior minus average tip distance.**

<table>
<thead>
<tr>
<th></th>
<th>1-1 category</th>
<th>2-1 category</th>
<th>2-2 category</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>G1c</td>
<td>G2</td>
<td>G7a</td>
</tr>
<tr>
<td>Dc</td>
<td>257.85*</td>
<td>237.92</td>
<td>815.43‡</td>
</tr>
<tr>
<td>Dn</td>
<td>660.41</td>
<td>−40.95</td>
<td>1130.88‡</td>
</tr>
</tbody>
</table>

* Significantly small distance (P < 0.05).
‡ Significantly small distance (P < 0.01).
§ Significantly large distance (P < 0.01).


