

High DNA Sequence Variability at the $\alpha 1$ Na/K-ATPase Locus of *Artemia franciscana* (Brine Shrimp): Polymorphism in a Gene for Salt-Resistance in a Salt-Resistant Organism

Alberto G. Sáez,*† Ricardo Escalante,* and Leandro Sastre*

*Instituto de Investigaciones Biomédicas CSIC-UAM, Madrid, Spain; and †Alfred Wegener Institute, Bremerhaven, Germany

We previously reported that the Na/K-ATPase $\alpha 1$ subunit coding gene showed signs of being a very polymorphic locus in *Artemia franciscana*. This species is adapted to highly saline waters, and the Na/K-ATPase $\alpha 1$ isoform presumably plays a key role in this adaptation. Therefore, we were interested in further study of the $\alpha 1$ Na/K-ATPase polymorphisms to examine whether they might be due to an adaptation to salt resistance driven by natural selection. Using coding sequences from 10 genomic clones and 3 cDNAs, we observed that most substitutions are in synonymous positions (88.8%). The 12 nonsynonymous substitutions code for conservative amino acid replacements with an apparent scattered distribution across functional domains of the protein. Interspecific comparison between these sequences and two genomic clones from *Artemia parthenogenetica* containing 1,122 bp of the $\alpha 1$ Na/K-ATPase locus coding sequence showed independence of the synonymous/nonsynonymous ratio in the comparison within *A. franciscana* and between *A. franciscana* and *A. parthenogenetica*, which fits the neutral model of evolution. Since there were no previous studies on DNA polymorphism for other *A. franciscana* genes, we also studied variability at the *Actin 302* locus for comparison. Both loci were amplified by reverse transcription-polymerase chain reaction, and 20 sequences were obtained for each. This study shows that the amplified region of the $\alpha 1$ Na/K-ATPase gene is 3.5 times as polymorphic as the *Actin 302* gene and 2.9 times as heterozygotic. Interestingly, under a model of neutral evolution, the data observed would be expected with a probability of approximately 0.05, suggesting an excess of intraspecific variation of $\alpha 1$ Na/K-ATPase with respect to *Actin 302*. Restriction fragment length polymorphism studies show similar patterns of polymorphism along the ~41-kb span of the $\alpha 1$ Na/K-ATPase locus. Most of the nucleotide differences are linked in a few haplotypes, although recombination events are also inferred from the data. We propose a possible explanation for the high polymorphic levels at the $\alpha 1$ Na/K-ATPase locus which invokes positive selection acting tightly to the locus in transiently isolated or semi-isolated subpopulations.

Introduction

Among metazoans, the brine shrimp (*Artemia* spp.) are the organisms that tolerate the widest range of environmental salt concentrations and the most diverse chemical composition of the medium (Browne 1992). Survival under such extreme conditions is possible because, among other adaptations, *Artemia* have specialized organs that extrude salt from the isosmotic internal hemolymph to the external medium. In early developmental stages, the larva secretes the salts mainly through the salt gland (dorsal organ in the cephalothorax) and midgut (Conte 1984). For the same purpose, the adult uses glands located in the metaepipodites of the thoracic appendages, the maxillary glands, and the midgut (Holliday, Roye, and Roer 1990). High levels of Na/K-ATPase activity have been found in all of these organs, and this enzymatic activity increases in proportion to the external salinity of the medium (Cortas et al. 1989; Holliday, Roye, and Roer 1990). Consequently, and given the function of this protein in Na⁺ and K⁺ transport through the plasma membrane (Vasilets and Schwartz 1993), the Na/K-ATPase has been proposed to play an essential role in the extraordinary osmotic resistance of *Artemia* (Cortas et al. 1989; Conte 1989; Holliday, Roye, and Roer 1990).

Key words: *Artemia franciscana*, osmoregulation, population genetics, molecular evolution, Na/K ATPase, neutral evolution.

Address for correspondence and reprints: Leandro Sastre, Instituto de Investigaciones Biomédicas, Arturo Duperier, 4, 28029 Madrid, Spain. E-mail: lsastre@biomed.iib.uam.es.

Mol. Biol. Evol. 17(2):235–250. 2000

© 2000 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

Na/K-ATPase is a heterodimeric protein composed of two subunits: an α subunit that is homologous to P-type ATPases and contains most of the active centers, and a smaller β subunit (reviewed by Lingrel et al. 1990). Two genes coding for different α subunit isoforms have been cloned from *Artemia franciscana*, $\alpha 1$ (cDNA 136; Macías, Palmero, and Sastre 1991) and $\alpha 2$ (Baxter-Lowe et al. 1989). A gene coding for the β subunit has also been cloned (Bhattacharyya, Bergstrom, and Hokin 1990). The proteins encoded by the $\alpha 1$ Na/K-ATPase and $\alpha 2$ Na/K-ATPase genes are 73.8% identical in their deduced amino acid sequences (Macías, Palmero, and Sastre 1991). The $\alpha 1$ form is expressed in all of the proposed osmoregulatory organs of the larvae: the salt gland, the midgut, and the antennal gland (Escalante, García-Sáez, and Sastre 1995). The level of $\alpha 1$ mRNA accumulation increases from cysts to larvae, when the resumption of development takes place after cryptobiosis (Escalante et al. 1994). The other gene, $\alpha 2$, shows restricted expression to the salt gland in early larval stages (Guo and Hokin 1989; Sun et al. 1992; Escalante et al. 1994; Escalante, García-Sáez, and Sastre 1995). The broader temporal and spatial pattern of expression of the $\alpha 1$ Na/K-ATPase gene suggests that the corresponding $\alpha 1$ protein isoform is the main factor responsible for the osmoregulatory role of Na/K-ATPase in *A. franciscana*.

In an attempt to characterize the transcriptional regulation and the structure of this gene, genomic clones of $\alpha 1$ were isolated from *A. franciscana*. The length (approximately 41 kb) and exon-intron distribution (14

introns, 10 of them shared with mammals) were described previously (García-Sáez, Perona, and Sastre 1997). An unexpected degree of heterogeneity was found in the restriction maps of the genomic clones and in their exon coding sequences, without parallel in other genes whose exon-intron structures have been characterized in this organism (Escalante and Sastre 1994; Ortega, Díaz-Guerra, and Sastre 1996; Sastre 1999). The pairwise differences were of 0.4%–3.5% in translated regions and of 1.6%–8.2% in untranslated regions. There were also differences in the promoter regions of the genes. Based on these differences, two different clones were isolated; they showed 6% divergence in the nucleotide sequences, as well as different promoter activities when assayed in transfection experiments in cell culture. The introns were not sequenced, but their sizes also varied in some of the genomic clones isolated. Synonymous and nonsynonymous substitutions, insertions/deletions, and tandem duplications were found among the differences. It was shown, however, that the Na/K-ATPase $\alpha 1$ subunit is encoded by a single gene in *A. franciscana*, since restriction pattern analysis of DNAs from different individuals presented one or two fragments that hybridized to each $\alpha 1$ probe, in agreement with the expected distribution for different alleles of a single gene (García-Sáez, Perona, and Sastre 1997).

In the present study, we investigated the polymorphism of the *A. franciscana* $\alpha 1$ Na/K-ATPase gene (hereafter called $\alpha 1$) in greater detail. We used Southern blotting to verify the broad degree of variation at the locus suggested by the genomic clones. Sequences of the previously described cDNA and genomic clones (Macías, Palmero, and Sastre 1991; García-Sáez, Perona, and Sastre 1997), as well as new cDNA and reverse transcription–polymerase chain reaction (RT-PCR) sequences, were used to assess the degree and distribution of intraspecific variability at the coding positions of the $\alpha 1$ gene. Given the putatively essential role of the Na/K-ATPase $\alpha 1$ subunit in the adaptation of *Artemia* populations to highly saline environments, it is tempting to relate the intraspecific variability of $\alpha 1$ in *A. franciscana* to the effect of natural selection. We used several statistical tests, for which the $\alpha 1$ gene from *Artemia parthenogenetica* was partially cloned and sequenced, to study whether there are departures in the distribution of DNA sequence variation from neutral evolution (Kimura 1983). Since there were insufficient studies on nucleotide polymorphism at other *A. franciscana* loci for comparison, we also sequenced and analyzed the *Actin 302* locus (Macías and Sastre 1990) for intra- and interspecific DNA sequence variation.

Materials and Methods

Origin of *A. franciscana* Samples

The *Artemia* cysts (cryptobiotic gastrula stage embryos) used in restriction fragment length polymorphism (RFLP) or RT-PCR experiments were obtained from the same source as those used to synthesize the libraries from which cDNA and the genomic clones were isolated (lot 3556, San Francisco Bay Brand Co., San Francisco,

Calif.). The provider assured us that cysts of this lot were collected in the San Francisco Bay (Calif.). RFLP analyses using ribosomal probes showed that the pattern obtained from lot 3556 was the same as that previously observed for other *A. franciscana* samples from the San Francisco Bay and differed from that for specimens from the Great Salt Lake (Utah) (J. Cruces, personal communication). Cysts were cultured at 30°C to activate their embryonic development so that one nauplius hatched from each cyst. The size of a nauplius (first larval stage), from which RNA or DNA was extracted, is approximately 0.5 mm; it was consequently necessary to use more than one individual for the extractions. Since inbred *Artemia* lines are not available, it was not possible to obtain genetically homogeneous sets of *A. franciscana* for the RFLP or the RT-PCR experiments, and so aleatory pools of animals were used.

RFLP Studies

Cysts were cultured for 20 h at 30°C in 0.25 M NaCl as previously described (Osuna and Sebastián 1980), and DNA was extracted from the emerged nauplii (Cruces, Sebastián, and Renart 1981). DNA was digested with *EcoRI*, *HindIII*, and the combination of both. Aliquots of 15 μ g of DNA per lane were electrophoresed in 0.8% agarose gels, which were then transferred under alkaline conditions (0.4 N NaOH) to nylon membranes (Zeta-Probe, Bio-Rad, Richmond, Calif.). These membranes were prehybridized for 2 h in 7% SDS, 500 mM NaPO₄ (pH 7.2), and 1 mM EDTA (pH 7.0) at 65°C (Church and Gilbert 1984). Hybridization was performed for 15 h at 65°C with 10⁶ cpm/ml of probe in the same buffer. Washes were repeated three times for 30 min each in highly stringent conditions (0.1 \times SSC, 1% SDS at 65°C) or, when indicated, at low stringency (2 \times SSC, 1% SDS at 65°C). Under these conditions, no cross-hybridization with the $\alpha 2$ gene was detected. The locations and approximate lengths of the $\alpha 1$ probes are shown in figure 1. Probe E is the *HindIII-HindIII* fragment from clone gArATCa23, which contains most of exon 7 from the *sarco/endoplasmic reticulum Ca-ATPase* gene (Escalante and Sastre 1994).

Isolation and Characterization of the cDNA (*A. franciscana*) and Genomic Clones (*A. parthenogenetica*)

cDNA clones were isolated from an *A. franciscana* cDNA library inserted in the *Bam*HI site of the pUC18 plasmid vector using a partial cDNA clone as probe, as previously described (Macías, Palmero, and Sastre 1991). Genomic clones were isolated from a library constructed in the EMBL-3 lambda vector using DNA from a cyst population composed mainly of *A. parthenogenetica* (Franco et al. 1994). Clones from this library (7.5 \times 10⁵) were transferred to nitrocellulose filters and hybridized to the labeled insert of cDNA clone 136 (Macías, Palmero, and Sastre 1991; accession number X56650) in 6 \times SSC, 0.5% SDS, 5 \times Denhardt's solution, 100 μ g/ml of calf thymus DNA, and 10⁶ cpm/ml of the probe overnight at 65°C. Filters were washed three times in 0.1 \times SSC, 0.1% SDS for 30 min at 65°C.

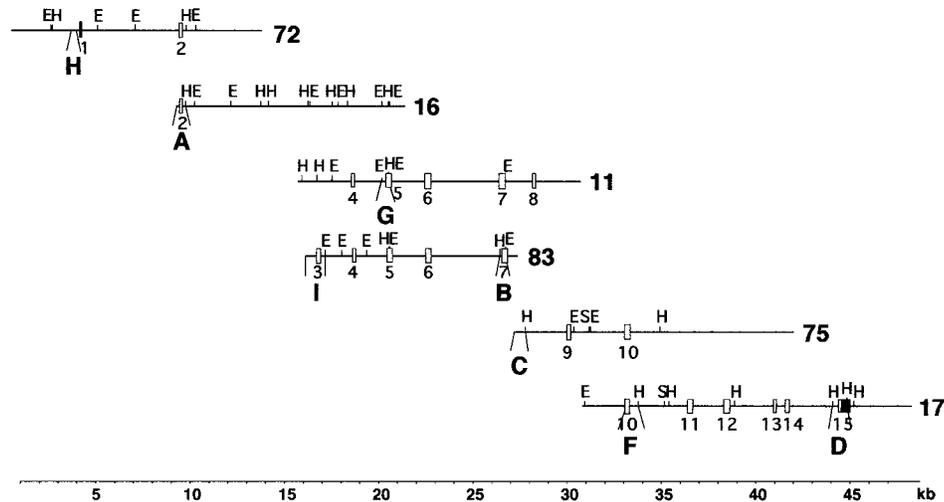


FIG. 1.—Locations of the eight probes used in the restriction fragment length polymorphism studies on the restriction maps of the $\alpha 1$ Na/K-ATPase genomic clones from which they were isolated. Vertical lines below the corresponding clones indicate the positions of probes A–D and F–I used in the Southern blots for *Artemia franciscana* genomic DNA (fig. 2 and data not shown). The indication of the probes is overlapped to figure 1 from García-Sáez, Perona, and Sastre (1997), which describes the structure of the $\alpha 1$ Na/K-ATPase genomic clones isolated. The clones are aligned by the positions of the 15 exons of the gene, shown as open boxes (coding) or black boxes (noncoding). EcoRI, HindIII, and SalI restriction sites are indicated by E, H, and S, respectively.

The clone restriction map was determined after single and double digestions with the enzymes indicated. Restriction fragments of the genomic clones containing coding sequences were identified by hybridization with cDNA clone 136 and cloned in the pUC18 vector. In some cases, smaller fragments were made by digestion with *TaqI* or *HinfI* to establish the nucleotide sequences of the coding regions. The location of each exon on the large restriction fragments was deduced from nucleotide sequence data or from PCR experiments using primers that anneal in or close to the exons and universal sequencing primers that anneal close to the pUC18 cloning site.

The cysts used to construct the *A. parthenogenetica* genomic library contained a small proportion of *A. franciscana* animals (Ortega, Díaz-Guerra, and Sastre 1996). For this reason, the species from which genomic clones Par3 and Par7 originated was confirmed by PCR using DNA from pure *A. franciscana* and *A. parthenogenetica* populations. The differences in intron regions of the genomic clones were used to design oligonucleotides specific for genomic clones Par3 or Par7 (5'-AGATGGCAAAGGCAAGC-3' and 5'-GGGACTCAGATAAAGGAGAG-3') or for the previously characterized *A. franciscana* genomic clones (5'-AGATGGCAAAGGCAAGC-3' and 5'-GGTACTGAGATACAGCAGAG-3'). The Par3- and Par7-specific oligonucleotides amplified DNA fragments of the predicted size from pure *A. parthenogenetica* DNA but not from *A. franciscana* DNA. Converse results were obtained with the oligonucleotides derived from *A. franciscana* genomic clones: amplification products were obtained from *A. franciscana* DNA but not from *A. parthenogenetica* DNA (data not shown).

Amplification of cDNA Fragments by RT-PCR

Cysts were cultured as above, and aliquots of more than 10⁵ nauplii were obtained. RNA was purified from

each aliquot with Trizol (GIBCO-Life Technologies, Gaithersburg, Md.). Five micrograms from each RNA aliquot were used as substrate in reverse transcriptase (RT) reactions using specific primers for $\alpha 1$ or *Actin 302*. One fifth of the RT reaction was used as substrate for PCR amplification. Ten independent RT-PCR reactions were performed for $\alpha 1$, and one to three clones obtained from each reaction were sequenced. Five different RT-PCR reactions were also carried out for *Actin 302*, and four clones from each reaction were sequenced. The low-error-rate Expand High Fidelity PCR System (Boehringer Mannheim, Germany) was used for these amplifications. PCR products were purified in agarose gels and ligated to the pGEMT easy vector (Promega Corporation, Madison, Wis.). The oligonucleotides used for RT and PCR priming were 5'-ATGGCAAAGGCAAGCAAAG-3' and 5'-GTTGGCAGAGTTGAATGG-3' for $\alpha 1$ Na/K-ATPase and 5'-GTTACTTCTTTGATTGAGGCTCG-3' and 5'-CGGGCAAGTCATTGAATG-3' for *Actin 302*.

DNA Sequencing

The nucleotide sequences of both strands of the cDNA clones were determined using pUC/M13 universal sequencing primers after the generation of progressive deletions from both ends of each clone (Guo, Yang, and Wu 1983). The coding-region sequences of the genomic clones were determined for both DNA strands using universal sequencing primers or specific primers complementary to the previously determined sequences of the cDNA clones. Nucleotide sequences were determined by the dideoxy chain-termination method (Sanger, Nicklen, and Coulson 1977) as modified by Chen and Seeburg (1985) for sequencing double-stranded DNA. The nucleotide sequences of the RT-PCR-amplified fragments were determined for one strand with the Taq Dye Deoxy Terminator Cycle Sequencing Kit and

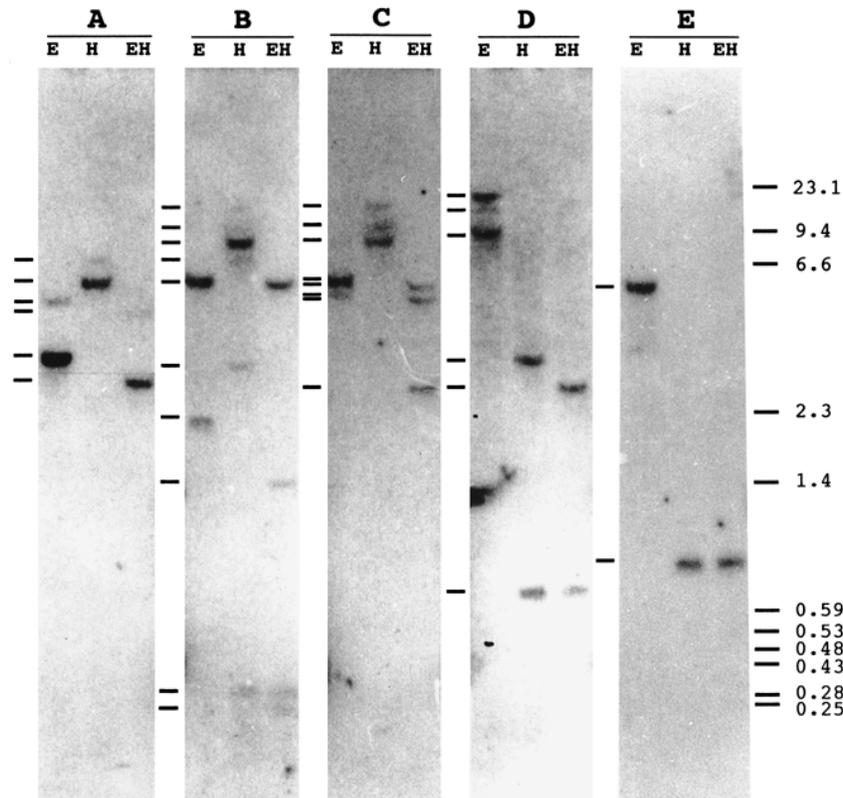


FIG. 2.—Southern blot hybridization of *Artemia franciscana* genomic DNA. DNA obtained from *A. franciscana* cysts (lot 3556, San Francisco Bay brand) was digested with *Eco*RI (E) or *Hind*III (H) or was double-digested with both enzymes (EH). Panels A–D were hybridized to probes A–D (specific to $\alpha 1$ *Na/K-ATPase*; fig. 1). One of these membranes was stripped of the probe and rehybridized to probe E, from the sarco/endoplasmic reticulum *Ca-ATPase* (Escalante and Sastre 1994). Dashes on the left of the panels indicate the positions of the fragments that hybridize to each probe. On the right, the mobility of the marker used is shown in kilobases.

a 373A Sequencer (Applied Biosystems). All sequences were independently incorporated into the alignments. The differences detected were then visually confirmed in the chromatograms.

Statistical Analyses

DnaSP versions 2.82 and 2.90 (Rozas and Rozas 1997) were used to calculate statistics of intra- (π , θ , etc.) and interspecific differentiation (divergence) and the number of synonymous and nonsynonymous sites (Nei and Gojobori 1986, eqs. 1–3) and to perform the neutrality tests (Hudson, Kreitman, and Aguadé 1987; Tajima 1989; Fu and Li 1993). Only total number of segregation sites was considered, not total number of mutations. For the analyses of RT-PCR-amplified regions, for $\alpha 1$, we used the following samples: $\alpha 1$ –21, composed of the colinear 20 RT-PCR clones plus the homologous region from cDNA 136; $\alpha 1$ –21+102+83, composed of $\alpha 1$ –21 plus the homologous available regions from cDNA 102 and genomic clone 83 (only partially overlapping among themselves and with the RT-PCR clones); $\alpha 1$ –21 – (2+20), which is not a random sample but $\alpha 1$ –21 without clones 2 and 20 (the most differentiated RT-PCR clones) to study the effect of their elimination from $\alpha 1$ –21. For *Actin 302*, we used Act-21, made of the 20 RT-PCR clones plus the homologous region from actin cDNA 302. Sequence alignments are available at asaesz@awi-bremerhaven.de.

Results

$\alpha 1$ *Na/K-ATPase* RFLPs

Our first question was that of how the DNA variability found in the $\alpha 1$ genomic clones (García-Sáez, Perona, and Sastre 1997) was distributed throughout the locus. We wanted to determine the number of restriction fragments detected by Southern blot, their relative intensity, and whether or not the number of polymorphic bands was similar throughout the locus (40–45 kb). Eight probes subcloned from the genomic clones were used (fig. 1). *Artemia franciscana* DNA was digested with *Eco*RI, with *Hind*III, or simultaneously with both restriction enzymes. Four different blots were hybridized to probes A–D (see fig. 1) under high-stringency conditions (fig. 2). As a control for complete DNA digestion and for comparison with another locus, one hybridized membrane was stripped and rehybridized to a fragment from the *sarco/endoplasmic reticulum Ca-ATPase* (probe E). This control probe revealed one band in each digestion corresponding to the expected size according to the restriction maps of the genomic clones (Escalante and Sastre 1994). However, the $\alpha 1$ probes A–D hybridized to several bands in each digestion. In the double digestions with *Eco*RI and *Hind*III (fig. 2), probe A revealed two bands (of 4.3 and 2.8 kb), probe B revealed four (5.9, 1.7, 0.45, and 0.30 kb), probe C revealed three (5.8, 5, and 2.7 kb), and probe D revealed two (2.7 and

0.8 kb). Interestingly, one of the restriction fragments with each probe presented a strong signal, while the other fragments were much weaker. Some of them had corresponding restriction fragments in the isolated genomic clones (fig. 1), but most of them did not (7 out of 11). For example, the 0.8-kb fragment obtained with probe D corresponds to a *HindIII-HindIII* segment in clones 13, 81, and 17, but the larger 2.7 kb band was absent from the genomic clones. Reciprocally, some of the restriction fragments in the genomic clones were not observed in the Southern blots; for instance, the 5.7-kb *HindIII-EcoRI* fragment from clone g36 was not detected with probe H (data not shown). Probes F, G, H, and I revealed a similar pattern of variability (data not shown).

$\alpha 1$ Variability Based on cDNA and Genomic Clone Sequences

The initial screening of an *A. franciscana* cDNA library for the Na/K-ATPase $\alpha 1$ subunit yielded several clones. Only one, cDNA 136, contained the complete coding region and was previously described (Macías, Palmero, and Sastre 1991). The unexpected variability found later in the genomic clones (García-Sáez, Perona, and Sastre 1997) raised our interest in the other cDNA clones that had also been isolated: c102 and c130 ("c" or "g" are used to indicate cDNA or genomic clones, respectively). The three cDNA clones (c136, c102, and c130) were obtained without any consideration of their variability; they thus form a small but random sample that may provide information about the levels and patterns of nucleotide substitutions at $\alpha 1$. Figure 3A shows their physical maps. Two restriction site polymorphisms are shown in the coding region (*KpnI* and *BglII*), and one is shown in the 3' noncoding region (*HindIII*). The complete sequences for c102 and c130 were obtained (EMBL accession numbers AJ389884 and AJ389885). The polymorphic differences found in the coding region with respect to c136 or with respect to the sequences from the genomic clones are shown in figure 4. The numbers and percentages of substitutions among the coding regions of these clones are given in table 1. The values observed range from 0% to 4.0%. The two most different cDNA clones are c136 and c130 (1.86% differences). On average (total number of pairwise differences over total number of sites compared) these clones differ in 1.35% of their positions. c130 and c136 were also compared in their 3' noncoding areas, where they differ in 4 indels and 36 substitutions over 500 positions (excluding indels, 7.2%).

The isolation of the genomic clones for $\alpha 1$ rendered 15 clones, 10 of which were totally or partially sequenced in their coding exons (García-Sáez, Perona, and Sastre 1997). The sampling here cannot be considered random, as clones with very similar restriction maps tended to be skipped for sequencing. Comparison of sequences from genomic and cDNA clones can nevertheless provide insights into the variability of the $\alpha 1$ locus. In particular, it can be informative by showing the ratio of nonsynonymous versus synonymous polymorphisms and the types of amino acid replacements. The

areas sequenced for each clone are shown in figure 3C. Many do not overlap, since the lambda clones used to build the library have an average insertion size of 15 kb, but the gene is approximately three times this length. Figure 4 shows the nucleotide differences found between the cDNA and the genomic clones, and table 1 shows the numbers and percentages of pairwise differences. Some of the clones were identical in their shared sequenced parts, such as g13, g81, and c130, but most presented differences. The most distant were g83 and c130 (4% differences).

Most of the differences observed are silent. Of 107 polymorphisms found among the genomic and cDNA clones (104 segregating sites), 12 imply amino acid replacements (11.2%) (fig. 4). Figure 5 shows the deduced amino acid sequences of the clones. The protein polymorphisms are conservative (e.g., Leu/Ile, Glu/Asp, Ser/Asn, and Cys/Ser), although their functional relevance cannot be excluded. To investigate whether the amino acid substitutions are restricted to particular domains of the protein, we have depicted the location of the putative transmembrane or catalytic domains of the Na/K-ATPase α subunit (Serrano 1989) (fig. 5). There are 10 proposed transmembrane regions (H1–H10); regions a–j are considered important functional domains, as they are conserved among all P-type ATPases. The functional relevance of many of these domains has also been confirmed by mutagenesis experiments (Andersen and Vilsen 1995). The replacements are apparently randomly distributed throughout the protein sequence, although just four of them (two marginally) are found in the proposed domains (which account for 47% of the protein sequence). To our knowledge, none of the replaced positions has been found to be critical for Na/K-ATPase activity in the extensive point mutation studies conducted so far (reviewed by Vasilets and Schwartz 1993; Lingrel and Kuntzweiler 1994). We notice, however, that one of the replaced amino acid sites (V/A in H4) has been reported to have functional effects in the homologous sarco/endoplasmic reticulum Ca-ATPase (Clarke et al. 1993).

Polymorphism Versus Divergence in Comparing cDNA and Genomic Clones from *A. franciscana* with Genomic Clones from *A. parthenogenetica*

We wanted to compare the intraspecific polymorphisms described in $\alpha 1$ of *A. franciscana* with its interspecific differences from *A. parthenogenetica*, a closely related species. The isolation of two $\alpha 1$ genomic clones from *A. parthenogenetica* allowed us to obtain 1,122 bp homologous to *A. franciscana*'s exons 2–4 and 6–8 (exon 7 is incomplete; see appendix), with identical exon-intron borders (EMBL accession numbers AJ269674–AJ269678 and AJ269679–AJ269681). The two overlapping clones, named Par3 and Par 7, are shown in figure 3B. Exon 5 has probably been deleted from Par7, possibly due to the large size of the insert contained in the lambda clone (>18 kb). Exon 6 is present and is identical in both clones.

We were particularly interested in comparing the ratio of synonymous versus replacement substitutions

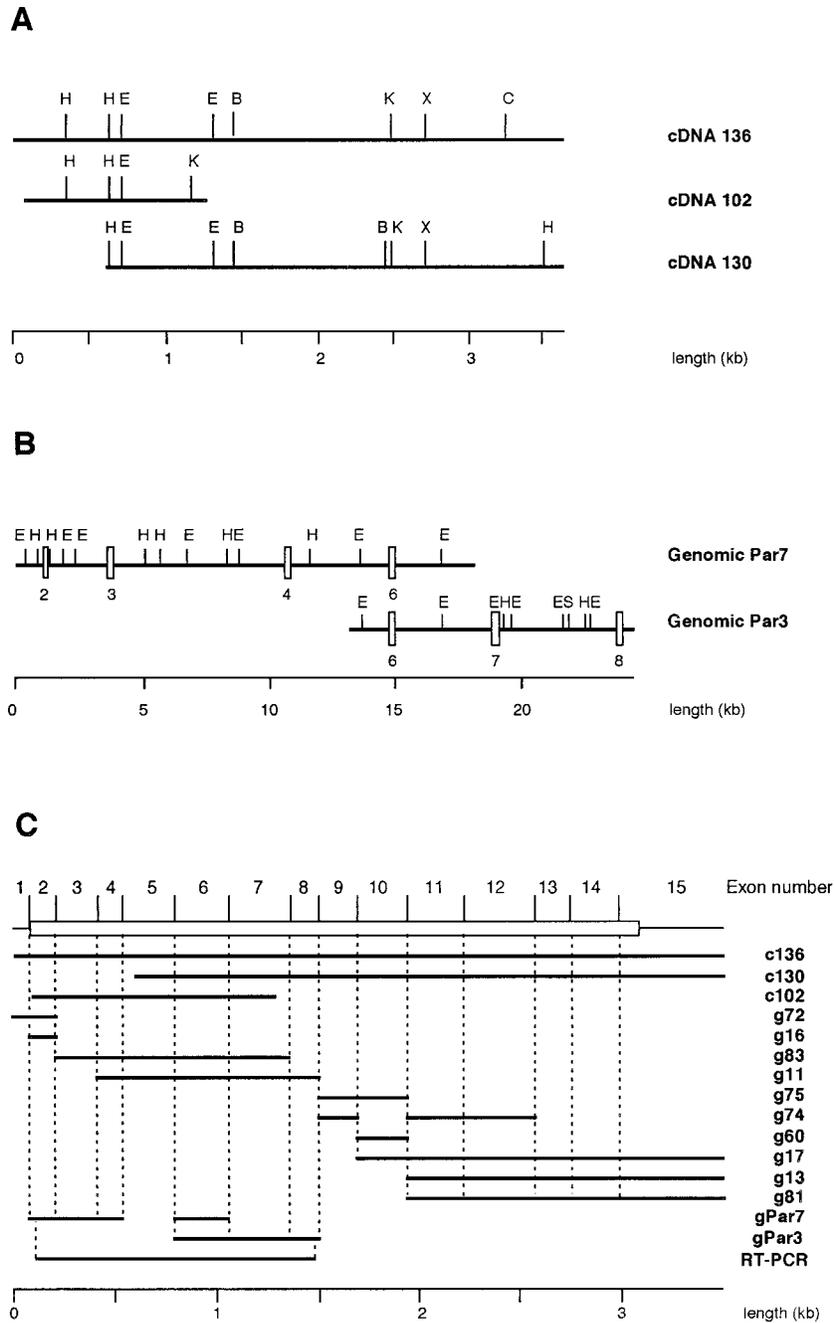


FIG. 3.—Restriction maps of the *Artemia a1 Na/K-ATPase* cDNA and genomic clones isolated. *A*, *Artemia franciscana* cDNA clones cDNA 102 and cDNA 130 were isolated from the same library as the previously characterized cDNA clone pArATNa136 (cDNA 136) (Macías, Palmero, and Sastre 1991). (*B*: *Bgl*II; *C*: *Cla*I; *E*: *Eco*RI; *H*: *Hind*III; *K*: *Kpn*I; *X*: *Xho*I). *B*, The genomic clones (Par3 and Par7) were isolated from an *Artemia parthenogenetica* library. The restriction maps were determined by single and double digestion with the enzymes *Eco*RI (*E*), *Hind*III (*H*), and *Sal*I (*S*). Boxes on the restriction maps represent the positions of the exons whose numbers are indicated underneath. *C*, Diagram showing the length and location of the coding regions sequenced in the cDNA, genomic, and RT-PCR clones compared in this study. Each horizontal bar represents the sequence of one clone, indicated on the right. Vertical dashed lines indicate the limits between exons, numbered at the top.

within *A. franciscana* and between *A. franciscana* and *A. parthenogenetica*. Different ratio for the within- and between-species comparisons would reveal nonneutral evolution for *a1*, suggesting a tendency to accumulate either replacement or synonymous substitutions in one lineage relative to the other. An appropriate test to detect deviations from neutrality in the synonymous-versus-

nonsynonymous ratio when comparing intraspecific and interspecific data was described by McDonald and Kreitman (1991). This test was applied to the 1,122 positions shared by the sequences from both species. The differences between them are shown (fig. 4) with indications for each substitution of whether is fixed or polymorphic and whether it induces a silent or a replacement substi-

Table 1
Differences Between the Nucleotide Sequences of *Artemia franciscana* cDNA (c102, c130, c136) and Genomic (g72, g16, g83, g11, g75, g74, g60, g17, g13, and g81) Clones and *Artemia parthenogenetica* Genomic Clones gPar7 and gPar3 (Apa; These Clones Were Identical in Their Overlapping Regions)

Clone	c136	c130	c102	g72	g16	g83	g11	g75	g17	g13	g81
c130	46/2,479 (1.86)										
c102	4/1,199 (0.33)	9/685 (1.31)									
g72	0/123 (0)		0/104 (0)								
g16	0/123 (0)		0/104 (0)	0/123 (0)							
g83	39/1,127 (3.46)	30/744 (4.03)	38/1,068 (3.56)								
g11	4/985 (0.41)	6/783 (0.77)	7/794 (0.88)			24/826 (2.91)					
g75	6/442 (1.36)	0/442 (0)							7/249 (2.81)		
g74	3/508 (0.59)	7/508 (1.38)						2/193 (1.04)	8/315 (2.54)	9/315 (2.86)	9/305 (2.95)
g60	3/249 (1.20)	0/249 (0)						0/249 (0)	7/249 (2.81)		
g17	41/1,408 (2.91)	30/1,408 (2.13)									
g13	30/1,068 (2.81)	0/1,068 (0)							22/1,067 (2.06)		
g81	28/1,050 (2.67)	0/1,050 (0)							23/1,049 (2.19)	0/994 (0)	
Apa	54/1,226 (4.4)	36/785 (4.59)	41/971 (4.22)	3/122 (2.46)	3/122 (2.46)	47/858 (5.48)	29/521 (5.57)				

NOTE.—The percentage of nucleotide differences is indicated in parentheses for each comparison.

DNA Sequence Variability at $\alpha 1$ and *Actin 302* Based on 20 Colinear Fragments Obtained by RT-PCR

Random overlapping sequences are necessary in order to calculate standard measures of genetic variation, such as π or θ . π is the nucleotide diversity, which is the average number of nucleotide differences per site between two sequences (Nei 1987, eq. 10.5). θ is $4N\mu$ (four times the effective population size multiplied by the mutation rate per site, sequence, and generation), a parameter usually estimated from S (the number of polymorphic sites) and the sample size (Nei 1987, eq. 10.3). In practice, π is a measure of heterozygosity, and θ , when calculated from S , is a measure of polymorphism. These two values, among others, also allow the performance of neutrality tests other than the previously mentioned McDonald and Kreitman test, such as Tajima's (1989) test or the HKA test (Hudson, Kreitman, and Aguadé 1987). The tests of neutrality test null hypotheses based on assumptions of the neutral model (Kimura 1983). Determination of whether or not the DNA sequence variation at $\alpha 1$ presents a neutral distribution is of use, since it gives clues about the evolutionary forces that have acted on this gene.

Accordingly, the variation at $\alpha 1$ was further characterized by comparing the nucleotide sequences of 20 randomly chosen colinear fragments generated by RT-PCR (EMBL accession numbers AJ389864–AJ389883). The amplified region was 1,362 nt, excluding the annealing primer regions, which code for 454 amino acids beginning at residue 8 of the Na/K-ATPase $\alpha 1$ subunit. For comparison, a fragment of 1,148 nt containing the

complete coding region of the *Actin 302* gene plus 17 nt of the 3' untranslated region was also amplified, and 20 sequences were obtained (EMBL accession numbers AJ269567–AJ269586). The polymorphisms found at both the $\alpha 1$ and the *Actin 302* loci are shown in figure 6, and the parameters obtained after their analyses are shown in table 3. Within the 20 RT-PCR clones plus the corresponding fragment from cDNA 136 ($\alpha 1$ –21; for a description of the samples, see *Materials and Methods*), we found 54 segregating sites and 55 substitutions (site 902 has two independent substitutions). Since the length of the alignment is 1,362 bp, the number of polymorphisms per site is 0.0396. In contrast, for the *Actin 302* gene, sample set Act-21 (the 20 RT-PCR fragments plus c302) contains only 13 polymorphisms along 1,148 bp (0.0113 polymorphisms per site). When π values from the two loci are compared, the difference is smaller: $\alpha 1$ –21 has a π of 0.0049, and Act-21 has a π of 0.0027. According to these data, the $\alpha 1$ fragment is thus 3.5 times as polymorphic as the actin region, and 1.8 times as heterozygotic. The larger difference between both loci for the polymorphism (or θ) than for heterozygosity (or π) is due to the abundance of polymorphisms found in only one sample (singletons) in $\alpha 1$ –21 (45 singletons; 82% of the substitutions); Act-21 has eight singletons (61%). Interestingly, the excess of single substitutions in $\alpha 1$ derives from just two of the RT-PCR clones, 2 and 20. Without these two, the π for $\alpha 1$ is very similar (slightly lower) to that of *Actin 302* (see table 3), and the polymorphism rates are much closer between the two loci, although that for $\alpha 1$ is still higher (1.4 times

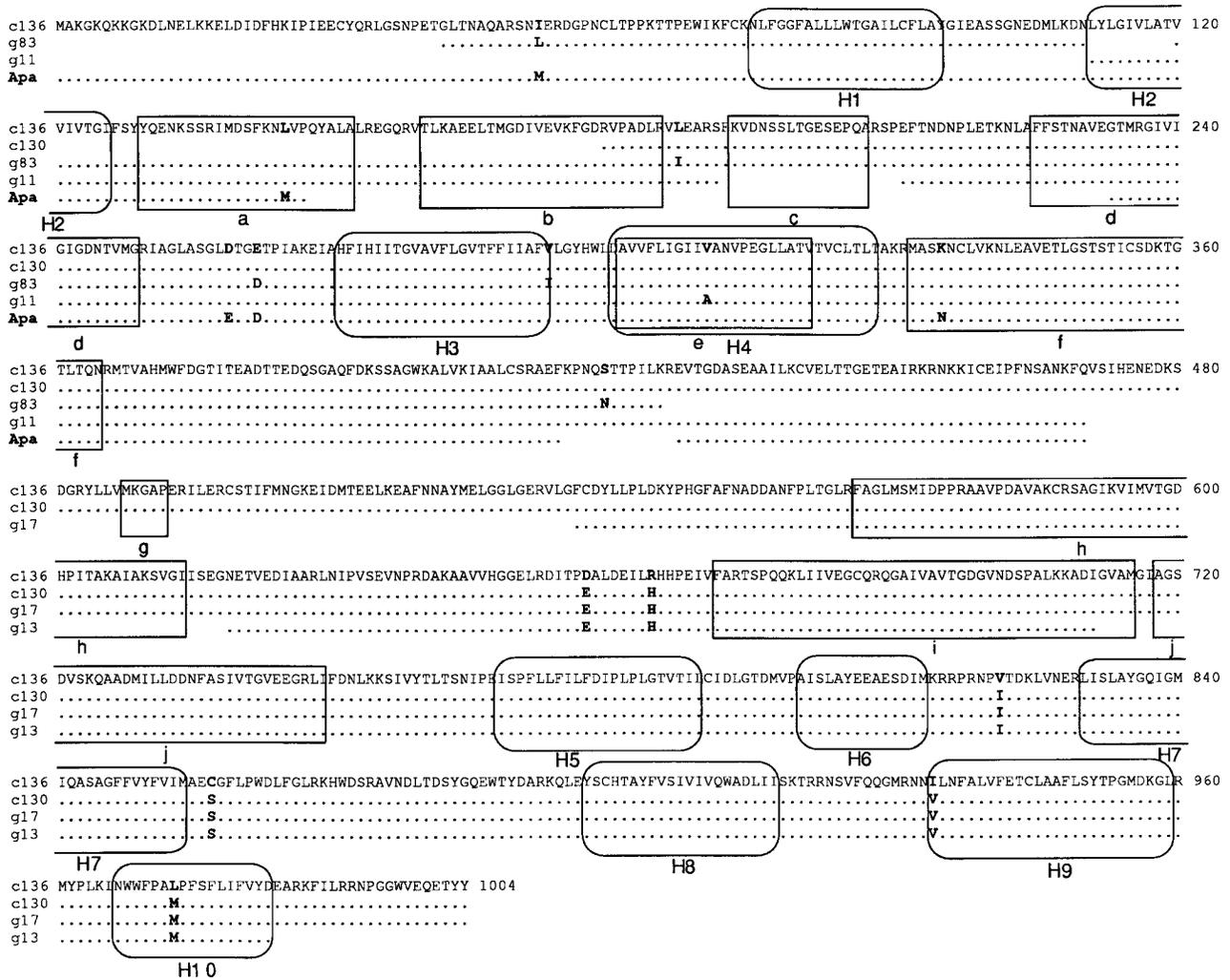


FIG. 5.—Differences in the deduced amino acid sequences from the *Artemia* $\alpha 1$ Na/K-ATPase clones. Upper lines show the amino acid sequence deduced from cDNA 136. The amino acid sequence of the other *Artemia franciscana* cDNA (c130) and genomic clones (g83, g11, g17, and g13) and *Artemia parthenogenetica* genomic clones (Apa) are shown underneath. Dots in the lower lines indicate the same amino acid as in c136. Square boxes (a–j) indicate protein regions that are conserved between P-type ATPases and have been shown to have functional relevance. Rounded boxes (H1–H10) are predicted transmembrane regions.

as high). These results are equally supported when only synonymous sites are considered (table 3), due to the relative abundance of synonymous versus nonsynonymous substitutions in both loci. The above-mentioned difference between π and θ is corroborated by Tajima's test, which detects significant deviations from the equation $\pi = \theta$, an equivalence that is predicted in the neutral model (Tajima 1989). If the corresponding D statistic is significantly different from zero, there is rejection of neutrality, as both values are significantly different; this is the case for the data set $\alpha 1$ –21.

The HKA test (Hudson, Kreitman, and Aguadé 1987) was also applied to these data. This is a conservative test that compares the ratios of intraspecific and interspecific variation between two or more loci. Under many assumptions based on neutral evolution, the null hypothesis predicts that both values covary because they are mainly due to the common forces of mutation rate and genetic drift. For the HKA tests, the RT-PCR-am-

plified regions of the $\alpha 1$ and *Actin 302* genes were used only for the calculation of the within-species variation; for between-species differences, we used the values shown in table 2 (these values apply to all available sites in the interspecific comparison for both loci). We found no departure from neutrality between $\alpha 1$ –21 and Act-21, considering synonymous sites only, or the total number of sites. Yet, the P values are low: 0.092 for all sites ($\chi^2 = 2.85$) and approximately the same value for synonymous sites; this is an indication of an excess of polymorphic sites in $\alpha 1$ compared with *Actin 302*.

The distribution of the polymorphisms in $\alpha 1$ –21 and Act-21 has been compared among exons (fig. 6). In the amplified regions from $\alpha 1$, exons 2 and 8 are incomplete; the actual lengths of these exons are 123 and 133 bp, respectively. There is an apparently heterogeneous distribution of the polymorphisms across the seven exons of $\alpha 1$ and the five exons of *Actin 302*. Two of the $\alpha 1$ exons, exons 4 and 8, present no polymorphism,

Table 2
Divergence Values for $\alpha 1$ Na/K-ATPase and Actin 302
Between *Artemia franciscana* and *Artemia parthenogenetica*

	<i>αNa/K-ATPase</i>	<i>Actin 302</i>
Total		
No. sites	1,122	1,002
No. substitutions	54	40
No. substitutions/no. sites	0.0481	0.0399
Synonymous		
No. sites	262	234.5
No. substitutions	49	39
No. substitutions/no. sites	0.1870	0.1663
Nonsynonymous		
No. sites	857	767.5
No. substitutions	5	1
No. substitutions/no. sites	0.0058	0.0013

NOTE.—Values obtained from all available homologous sites between both species shown in the appendix (*$\alpha 1$ Na/K-ATPase*) and from Ortega, Díaz-Guerra, and Sastre (1996) (*Actin 302*).

although their interspecific differences are not particularly low compared with the other exons. The statistical significance of this heterogeneity in *$\alpha 1$* exons has again been approached using the HKA test for the two more polymorphic exons, exons 2 and 6. All-site intraspecific variability in exon 2 has been compared with the intraspecific variability from the other six exons combined using the sample $\alpha 1-21$; the corresponding *P* value is 0.15. Similar analysis for exon 6 rendered a *P* value of 0.10. For synonymous sites, we obtained *P* values of 0.16 for exon 2 against the other six exons combined and 0.30 for exon 6 against all other exons from the RT-PCR fragments, respectively. The conclusion is that despite the heterogeneity in polymorphic levels among *$\alpha 1$* exons, the data do not show significant intraspecific differences among them when the HKA test is applied (i.e., taking into account the interspecific variation).

RT-PCR Sequences Compared with Genomic and cDNA Sequences

To integrate the sequence information generated from the RT-PCR reactions with the previously studied *$\alpha 1$* genomic or cDNA sequences, the four sequences with most overlap with the amplified region (cDNAs

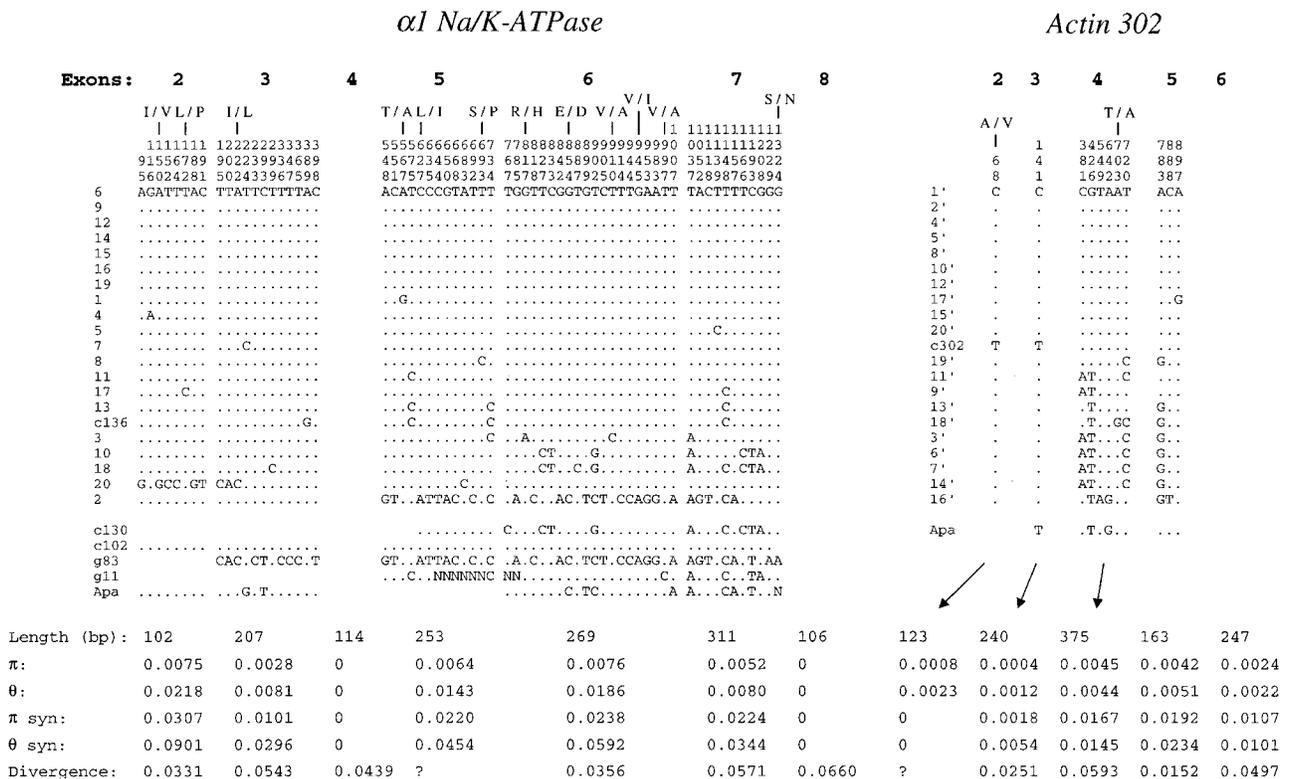


FIG. 6.—Nucleotide polymorphisms found at *$\alpha 1$ Na/K-ATPase* and *Actin 302* of *Artemia franciscana*. The differences among 20 clones obtained by RT-PCR are shown for 1,362 bp amplified for *$\alpha 1$* and 1,148 bp amplified for *Actin 302*. cDNA 136 for *$\alpha 1$* (c136) and cDNA 302 for *Actin 302* (c302) are included in the comparisons. The sequences have been ordered from common (top) to rare (bottom). A dot indicates that there is no substitution with respect to the nucleotide in the reference sequence at the top, numbered according to the cDNA 136 sequence (accession number X56650). The polymorphisms are separated in columns corresponding to the exons. The amino acid replacements are indicated above the respective sites. The total length, π , θ , π and θ for synonymous sites, and the interspecific divergence for each exon with respect to each set of 21 sequences are shown below. The polymorphisms for *$\alpha 1$* in cDNAs 130 and 102 (c130, c102) and in genomic clones 83 and 11 (g83, g11) are shown for comparison; empty spaces are positions that are absent in the corresponding clone. The bottom sequence lines in both loci show the corresponding nucleotides in *Artemia parthenogenetica* (Apa) for these polymorphic sites. N indicates missing positions, and question marks indicate unknown divergence values due to the unavailability of the corresponding *A. parthenogenetica* sequences.

Table 3
Values of Intra- and Interspecific Variability at $\alpha 1$ Na/K-ATPase and Actin302 in *Artemia franciscana* (vs. *A. parthenogenetica*)^a

Sample ^b	π^c	S^d	bp ^e	θ^f	Div. ^g	$\pi/\text{Div.}$	$\theta/\text{Div.}$	Hap. ^h	D^i	HKA: P^j	HKA: P_s^j	π syn ^k	S syn ^k	bp syn ^k	θ syn ^k
$\alpha 1-21$	0.005	54	1,362	0.0110	0.0484	0.0965	0.2247	15	-2.23*	0.092	0.091	0.0176	44	323.8	0.0386
$\alpha-21+102+83$. . .	0.008	53	1,068	0.0135	0.0466	0.1628	0.2923	15	-1.64	0.052	0.047	0.0284	45	259.0	0.0481
$\alpha-21-(2+20)$	0.003	21	1,362	0.0044	0.0478	0.0539	0.0947	13	-1.58	0.807	0.874	0.0093	16	323.8	0.0141
Act-21	0.003	13	1,148	0.0032	0.0419	0.0728	0.0793	12	-0.45	—	—	0.0110	11	269.3	0.0114

^a All of these values are only for the regions amplified in the RT-PCR experiment.
^b The composition of the samples is described in *Materials and Methods* ($\alpha-21-(2+20)$ is not a random sample).
^c Nucleotide diversity per site.
^d Number of polymorphic sites.
^e Length of every alignment excluding sites with missing information (there are no gaps).
^f Calculated as $\theta = S \sum 1/i$, from $i = 1$ to $i = n - 1$ (n is the number of samples).
^g Average proportion of nucleotide differences between species.
^h Number of haplotypes in each sample.
ⁱ Tajima's D statistic.
^j P value for an HKA test between the corresponding sample and Act-21, for all sites (P) and for synonymous sites (P_s).
^k π syn, S syn, bp syn, and θ syn are π , S , number of sites, and θ for the synonymous sites, respectively.
* $P < 0.01$.

130 and 102 and genomic clones 83 and 11; EMBL accession numbers AJ269664–AJ269668 and AJ269669–AJ269673) were included in the analyses shown in figure 6. Most of the substitutions in these clones are shared with the RT-PCR sequences. In particular, clones c130 and g11 are similar to RT-PCR sequences 18 and 10; clone g83 is similar to RT-PCR sequence 2 and to the 5' region of PCR sequence 20, and clone c102 is identical to the most commonly found RT-PCR haplotype. The $\alpha 1-21$ sample size was increased with the addition of the two longest genomic or cDNA clones, c102 and g83, and the Tajima and HKA tests were performed again. The resultant sample, $\alpha 1-21+102+83$, can still be considered a random sample, yet the decision about the number of clones to add to sample $\alpha 1-21$ is somewhat arbitrary. For $\alpha 1-21+102+83$, $\pi = 0.0078$ and $\theta = 0.0135$ (table 3). Both values are higher than those for $\alpha 1-21$, but the relative increase in π is greater than that in θ . This is due to the incorporation of clone g83 and its similarity to PCR sequence 2. The unequal increment in π and θ made Tajima's D appear closer to zero in $\alpha 1-21+102+83$ than in $\alpha 1-21$ (table 3); this is because the addition of g83 to $\alpha 1-21$ increases the average variability (π) more than the number of polymorphic sites (reflected in θ). The HKA test between sample $\alpha 1-21+102+83$ and sample Act-21 was also performed, as for the data set $\alpha 1-21$. The test rendered P values just above ($P = 0.052$; all sites) or below ($P = 0.047$; synonymous sites) the 5% level of significance of this conservative test. Again, as with the Tajima test for the data set $\alpha 1-21$, the nucleotide variation at the $\alpha 1$ locus shows signs of deviation from a strictly neutral origin.

Discussion

Levels of Intraspecific and Interspecific Nucleotide Sequence Variability

High levels of polymorphism at the $\alpha 1$ locus of *A. franciscana* were shown by the heterogeneity in the re-

striction maps and nucleotide sequences of the genomic clones that described the structure of the gene (García-Sáez, Perona, and Sastre 1997). In this study, we attempted to measure this variability in more detail and to test for its possible evolutionary significance. First, we compared the three cDNA clones isolated, c136, c102, and c130 (fig. 3A and C and table 1). The average percentage nucleotide sequence difference among them is 1.35%, very similar to the percentage found between the genomic clones (García-Sáez, Perona, and Sastre 1997). In contrast, the differences found in other *A. franciscana* loci for which one cDNA and one genomic coding sequence have been compared are much lower: 0.23% for *sarco/endoplasmic reticulum Ca-ATPase*, a gene of the same ATPase family and with similar structure to the gene studied here (Escalante and Sastre 1994); 0.1% for *Actin 302*; and 0.6% for *Actin 403* (Ortega, Díaz-Guerra, and Sastre 1996). The number of sequences compared for each locus is evidently small, but none of these other loci showed detectable levels of polymorphism in Southern blotting or according to the restriction maps of overlapping cDNA or genomic clones (Escalante and Sastre 1993, 1994; Ortega, Díaz-Guerra, and Sastre 1996; Sastre 1999). Since the same *A. franciscana* libraries were used in all of these previous studies, a different genetic origin of the clones can be excluded. Cloned genes from *A. franciscana* in other laboratories showed no evidence of variability either (e.g., Manzanares, Marco, and Garesse 1993; Franco et al. 1994).

The distribution of intraspecific variation at the $\alpha 1$ locus was studied by RFLP (fig. 2). Similar intraspecific variation levels were found throughout the entire $\alpha 1$ locus. For every probe assayed, we obtained a main hybridizing restriction fragment and one to three less abundant fragments. Also taking into account the results obtained on nucleotide sequence polymorphism, discussed later, these results could be interpreted by the existence of a majority of very similar alleles, with almost iden-

tical restriction maps, and a minor population of more divergent alleles.

We next attempted to quantify the amount of intraspecific variation at the *A. franciscana* $\alpha 1$ locus and to compare it with the interspecific variability. Genomic clones coding for approximately half of the gene were isolated from a close species, *A. parthenogenetica*. In addition, 20 independent, randomly generated RT-PCR clones coding for the same gene region were isolated and sequenced. Since there were no previous DNA variation studies for other *A. franciscana* loci, we also used the RT-PCR to amplify the *Actin 302* locus, since this was the only locus with available sequences for interspecific comparisons between *A. franciscana* and *A. parthenogenetica*.

The 20 amplified clones for $\alpha 1$ *Na/K-ATPase* showed a π value of 0.0049 ($\alpha 1-21$), 1.8 times as high as that for *Actin 302* (0.0027); the value for θ was 0.0110 ($\alpha 1-21$), 3.4 times as high as that for Act-21 (0.0032). These results clearly demonstrate the higher variability of the $\alpha 1$ gene compared with that of the *Actin 302* gene. This variability seems to be mainly due to the existence of some divergent alleles, such as those represented by RT-PCR clones 2 and 20. If these two clones are not considered (data set $\alpha 21 - (2+20)$), the results obtained, $\pi = 0.003$ and $\theta = 0.0044$, are very similar to those of *Actin 302*. Interestingly, the sequences of the divergent clones are very similar to those of previously isolated cDNA or genomic clones, such as c102 or g83 (fig. 6). When these two clones are included in the comparison (from $\alpha 1-21$ to $\alpha 1-21+102+83$), the nucleotide diversity increases by 59%, although both samples can be considered random. The fact that clone g83 contains a large number of mutations increases π drastically in $\alpha 1-21+102+83$, although θ does not increase in proportion, as g83 shares many of the substitutions with RT-PCR sequences 2 and 20. The presence of a large percentage of differences in a few sequences makes π more dependent on sample size. It might be considered that the variability described at the $\alpha 1$ locus could be due to a contamination in our batch of *A. franciscana* cysts with other *Artemia* species. The main argument against this possibility is that we have found recombinant haplotypes between differentiated and common sequences (e.g., RT-PCR clones 2 and 20; fig. 6).

The sequence analyses of interspecific sequence variation showed a similar proportion of fixed differences for both genes, but a lower rate of amino acid replacement in *Actin 302* than in $\alpha 1$. This result is not unexpected, since actins are highly conserved proteins: 249 out of 375 amino acid residues were identical in 95% of more than 100 actins that were analyzed by Sheterline and Sparrow (1994). Nonetheless, $\alpha 1$ and *Actin 302* showed similar percentages of synonymous differences between the two *Artemia* species, albeit the *Actin 302* synonymous rate was slightly lower, suggesting similar mutation rates for both genes. Therefore, the higher polymorphism of $\alpha 1$ is not just derived from a higher mutation rate, a finding that is statistically corroborated by the HKA test.

Neutral and Nonneutral Features at $\alpha 1$

To test for departures from neutral evolution at the $\alpha 1$ locus, we performed three types of neutrality tests. We first analyzed the rate of replacement versus synonymous substitutions, comparing the ratio within *A. franciscana* to that between *A. franciscana* and *A. parthenogenetica* (McDonald and Kreitman 1991). No signs of nonneutrality were found. The sequences of the genomic and cDNA clones were used for this test, since they span a longer region than do the 20 RT-PCR derived sequences. The fact that these clones are not completely overlapping should not affect the results, as we do not measure the overall level of variation, but, rather, the levels of silent variation within the observed differences. Another advantage of using the genomic and cDNA clones compared with the RT-PCR clones is that with the former, there is no risk of considering *Taq* polymerase errors; since the RT-PCRs were performed on pools of animals of heterogeneous genetic composition, we have no means of checking *Taq* polymerase errors, although a low-error-rate polymerase system was used. The hypothetical errors could affect the results of the McDonald and Kreitman test more than those of other tests, because nonsynonymous positions are in the majority. In any case, all RT-PCR clones were also tested and showed neutral behavior (data not shown).

The second neutrality test employed was the Tajima (1989) test. This test detected a significant deviation from the neutral expectation ($\pi = \theta$) in the data set $\alpha 1-21$ as a consequence of the high percentage of low-frequency substitutions in this sample (44 singleton sites of 54 segregating sites). When the $\alpha 1-21$ sample is augmented with clones c102 and g83, this significance is lost, as the similarity between clone g83 and RT-PCR clones 2 and 20 increases the frequency of many of the substitutions. We also used four similar tests from Fu and Li (1993) to analyze our data and obtained results very similar to those of the Tajima test: there is also rejection of neutrality for $\alpha 1-21$, although the *P* values are lower using Tajima's test (data not shown). Simulations to estimate the relative power of these tests by Simonsen, Churchill, and Aquadro (1995) showed better performance for Tajima tests than for Fu and Li tests, explaining the observed difference in *P* values. Both tests are considered useful for detecting selective sweeps or background selection, but we do not think that they reflect these events at $\alpha 1$. In such cases, a general deficiency of intraspecific variability would be found and, as discussed above, $\alpha 1$ has an excess rather than a deficiency of variability. The fact that the removal of sequences 2 and 20 from $\alpha 1-21$ transforms the statistics of variation of the remnant sample to values similar to those observed for data set Act-21 (table 3) indicates that haplotypes 2 and 20 represent the "excess" of intraspecific variation at $\alpha 1$ and not the "remainder" of a lost variation.

The HKA test applied to $\alpha 1$ and *Actin 302* also showed values that were significant or very close to significance (*P* = 0.05; table 3), again implying a surplus of polymorphism over the expectation of the neutral

model after taking interspecific values into consideration. There is an excess of synonymous (88.8%) over nonsynonymous substitutions (11.2%), an indication of selective constraints acting at $\alpha 1$. Also, nonsynonymous substitutions seem to be subjected to strong selective constraints, since the observed amino acid replacements are chemically conservative. The number of amino acid substitutions present in the original population could be much larger, as the number of overlapping sequences per site is quite small. Nonconservative amino acid changes could be taking place. It is also conceivable that some of the observed substitutions in nonsynonymous or even in synonymous positions offer selective advantage to their carriers, but we would expect most, even the nonsynonymous changes, to be neutral because of their conservative nature. In such a case, the forces increasing the presence of synonymous and nonsynonymous substitutions at $\alpha 1$ would be the same.

The existence in a population of several alleles maintained by natural selection produces an increase in the neutral intraspecific variation surrounding (or linked to) the balanced polymorphism sites (Strobeck 1983). A well-known case is that of *Adh* in *Drosophila melanogaster*. In the description of the HKA test, this locus was shown to present an abundance of polymorphisms when compared to its 5'-flanking region (Hudson, Kreitman, and Aguadé 1987). This excess of polymorphism was subsequently localized in a ~400-bp region whose only nonsynonymous substitution is contained in its middle part (Kreitman and Hudson 1991). If the excess of substitutions at *A. franciscana* $\alpha 1$ derives from the presence of one or more balanced polymorphisms at the locus or at a nearby one, we would expect an accumulation of substitutions around those polymorphisms. We have no candidates for such maintained polymorphisms, since none of the nonconservative changes appears to have a specific functional significance. The distribution of the substitutions among exons shows no clear sign of heterogeneity in intraspecific or interspecific comparisons, although an intriguing absence of polymorphisms was found in exons 4 and 8, two exons with interspecific substitution ratios close to and above the average value at $\alpha 1$, respectively.

An alternative explanation accounting for the polymorphic nature of $\alpha 1$ is that the population sampled in this study is derived from two or more previously isolated or semi-isolated subpopulations. It is not difficult to imagine how the subdivision of an *A. franciscana* population might take place: continental aquatic ecosystems, such as those in which *Artemia* lives, are more geologically dynamic than are terrestrial ecosystems, especially if they are coastal. In addition, it is known that *A. franciscana* habitats vary considerably in saline concentration or composition, not only between different ecosystems, but within a single one, both temporally and spatially (Lenz and Browne 1991). A simple geographical subdivision that would isolate populations from *A. franciscana*, later mixed in our sample, is unlikely to produce the observed results, as a similar differentiation pattern would have been expected for $\alpha 1$ and *Actin 302* (or other loci by RFLPs). Genetic isolation through a

suppression of the recombination between genomic regions closely linked to $\alpha 1$ is possible, but some evidence of recombination exists at positions 202–224 (RT-PCR clone 20), 398–548 (RT-PCR clone 2), and 1157–1166 (RT-PCR clone 2) (fig. 6). The existence of divergent haplotypes with a moderate degree of recombination among them suggest, rather, that populations divided geographically or ecologically in the past were subjected to environments that, at least in one subpopulation, imposed different selective pressure on the $\alpha 1$ locus. Environmentally driven positive selection acting specifically in one or more subpopulations can explain the abundance of synonymous substitutions that are strongly linked among themselves in a few haplotypes. During the hypothetical population subdivision, other loci would also be evolving independently between the subpopulations, but with no specific selective pressure acting on them, their evolution would be slower. A certain degree of differentiation would be expected, and, in fact, Act-21 shows a slight degree of linkage disequilibrium: 10 out of 78 pairwise comparisons show significant association (using the Fisher's exact test or the chi-square method).

In summary, the data obtained in this study show increased polymorphism at the $\alpha 1$ Na/K-ATPase locus (as compared with *Actin 302*) in brine shrimp *A. franciscana*. The analyses of these data indicate that the polymorphism observed is larger than expected for neutral evolution. Given the importance of Na/K-ATPase in osmotic regulation, the adaptation to the salinity conditions of the media could be the selective force acting on this gene to generate the polymorphism observed. Transient physical or ecological isolation of some subpopulations temporally exposed to different environmental conditions could have induced an accelerated positive selection process on the $\alpha 1$ locus of *A. franciscana* in at least one of the subpopulations, increasing the intraspecific levels of variability at the locus. However, a more extensive and intensive survey should be conducted on $\alpha 1$ and other genes from natural *Artemia* populations to advance the knowledge of the nature of the $\alpha 1$ polymorphism and to confirm or reject the proposed hypothesis.

Acknowledgments

We thank María Teresa Macías for her excellent technical assistance in sequencing the cDNA clones, Montserrat Aguadé for her invaluable help in many aspects of this work, Richard Hudson and Francisco Rodríguez-Trelles for their comments on a previous version of the manuscript, Julio Rozas for a test version of DnaSP, Paul Sharp for early initial advice, and Catherine Mark Tienmann for revision of the manuscript. We also thank Caro-Beth Stewart for her suggestions and the careful editing of the manuscript. This work was supported by grants from the Dirección General de Investigación Científica y Técnica (PB92-0076) and the Dirección General de Enseñanza Superior (PB95-0096). A.G.S. was partially supported by a postdoctoral grant from the EU to Linda Medlin (Alfred Wegener Institute).

LITERATURE CITED

- ANDERSEN, J. P., and B. VILSEN. 1995. Structure-function relationships of cation translocation by Ca- and Na,K-ATPases studied by site-directed mutagenesis. *FEBS Lett.* **359**:101–106.
- BAXTER-LOWE, L. A., J. Z. GUO, E. E. BERGSTROM, and L. E. HOKIN. 1989. Molecular cloning of the Na,K-ATPase α -subunit in developing brine shrimp and sequence comparison with higher organisms. *FEBS Lett.* **257**:181–187.
- BHATTACHARYA, K. K., E. E. BERGSTROM, and L. E. HOKIN. 1990. Molecular cloning of the β -subunit of the Na,K-ATPase in the brine shrimp, *Artemia*. The cDNA-derived amino acid sequence shows low homology with the β -subunit of vertebrates, except in the single transmembrane and the carboxy-terminal domains. *FEBS Lett.* **269**:233–238.
- BROWNE, R. A. 1992. Population genetics and ecology of *Artemia*: insights into parthenogenetic reproduction. *Trends Ecol. Evol.* **7**:232–237.
- CHEN, E. J., and P. H. SEEBURG. 1985. Supercoil sequencing: a fast simple method for sequencing plasmid DNA. *DNA* **4**:165–170.
- CHURCH, G. M., and W. GILBERT. 1984. Genome sequencing. *Proc. Natl. Acad. Sci. USA* **81**:1991–1995.
- CLARKE, D. M., T. W. LOO, W. J. RICE, J. P. ANDERSEN, B. VILSEN, and D. H. MACLENNAN. 1993. Functional consequences of alterations to hydrophobic amino acids located in the M4 transmembrane sector of the Ca-ATPase of sarcoplasmic reticulum. *J. Biol. Chem.* **268**:18359–18364.
- CONTE, F. P. 1984. Structure and function of the crustacean larval salt gland. *Int. Rev. Cytol.* **91**:45–106.
- . 1989. Molecular biology of larval osmoregulation. Pp. 371–376 in A. H. WARNER, T. H. MACRAE, and J. C. BAGSHAW, eds. *Cell and molecular biology of Artemia development*. Plenum Press, New York.
- CORTAS, N., M. ARNAOUT, J. SALON, and I. S. EDELMAN. 1989. Isoforms of Na,K-ATPase in *Artemia salina*: II. Tissue distribution and kinetic characterization. *J. Membr. Biol.* **108**:187–195.
- CRUCES, J., J. SEBASTIÁN, and J. RENART. 1981. Restriction mapping of the rRNA genes from *Artemia* larvae. *Biochem. Biophys. Res. Comm.* **98**:404–409.
- ESCALANTE, R., A. GARCÍA-SÁEZ, M.-A. ORTEGA, and L. SASTRE. 1994. Gene expression after resumption of development of *Artemia franciscana* cryptobiotic embryos. *Biochem. Cell Biol.* **72**:78–83.
- ESCALANTE, R., A. GARCÍA-SÁEZ, and L. SASTRE. 1995. In situ hybridization analyses of Na,K-ATPase α -subunits expression during early larval development of *Artemia franciscana*. *J. Histochem. Cytochem.* **43**:391–399.
- ESCALANTE, R., and L. SASTRE. 1993. Similar alternative splicing events generate two sarcoplasmic or endoplasmic reticulum Ca-ATPase isoforms in the crustacean *Artemia franciscana* and in vertebrates. *J. Biol. Chem.* **268**:14090–14095.
- . 1994. Structure of *Artemia franciscana* sarco/endoplasmic reticulum Ca-ATPase gene. *J. Biol. Chem.* **269**:13005–13012.
- FRANCO, F., M. M. BEHRENS, M. DÍAZ-GUERRA, and J. RENART. 1994. Structure and expression of a polyubiquitin gene from the crustacean *Artemia*. *Gene Expr.* **4**:19–28.
- FU, Y.-X., and W.-H. LI. 1993. Statistical tests of neutrality of mutations. *Genetics* **133**:693–709.
- GARCÍA-SÁEZ, A., R. PERONA, and L. SASTRE. 1997. Polymorphism and structure of the gene coding for the $\alpha 1$ subunit of the *Artemia franciscana* Na/K ATPase. *Biochem. J.* **321**:509–518.
- GUO, J. Z., and L. E. HOKIN. 1989. Differential gene expression of Na,K-ATPase α - and β -subunits in the developing brine shrimp, *Artemia*. Pp. 377–388 in A. H. WARNER, T. H. MACRAE, and J. C. BAGSHAW, eds. *Cell and molecular biology of Artemia development*. Plenum Press, New York.
- GUO, L.-H., R. C. A. YANG, and R. WU. 1983. An improved strategy for rapid direct sequencing of both strands of long DNA molecules cloned in a plasmid. *Nucleic Acid Res.* **11**:5521–5540.
- HOLLIDAY, C. W., D. B. ROYE, and R. D. ROER. 1990. Salinity-induced changes in branchial Na/K ATPase activity and transepithelial potential difference in the brine shrimp *Artemia salina*. *J. Exp. Biol.* **151**:279–296.
- HUDSON, R. R., M. KREITMAN, and M. AGUADÉ. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**:153–159.
- KIMURA, M. 1983. The neutral theory of molecular evolution. Cambridge University Press. Cambridge, England.
- KREITMAN, M., and R. R. HUDSON. 1991. Inferring the evolutionary histories of the *Adh* and *Adh-dup* loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. *Genetics* **127**:565–582.
- LENZ, P. H., and R. A. BROWNE. 1991. Ecology of *Artemia*. Pp. 237–253 in R. A. BROWNE, P. SORGELOOS, and C. N. A. TROTMAN, eds. *Artemia biology*. CRC Press, Boca Raton, Fla.
- LINGREL, J. B., and T. KUNTZWEILER. 1994. Na/K-ATPase. *J. Biol. Chem.* **269**:19659–19662.
- LINGREL, J. B., J. ORLOWSKI, M. M. SHULL, and E. M. PRICE. 1990. Molecular genetics of Na,K-ATPase. *Prog. Nucleic Acid Res. Mol. Biol.* **38**:37–89.
- MCDONALD, J. H., and M. KREITMAN. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**:652–654.
- MACÍAS, M.-T., I. PALMERO, and L. SASTRE. 1991. Cloning of a cDNA encoding an *Artemia franciscana* Na/K ATPase α -subunit. *Gene* **105**:197–204.
- MACÍAS, M.-T., and L. SASTRE. 1990. Molecular cloning and expression of four actin isoforms during *Artemia* development. *Nucleic Acids Res.* **18**:5219–5225.
- MANZANARES, M., R. MARCO, and R. GARESSE. 1993. Genomic organization and developmental pattern of expression of the engrailed gene from the brine shrimp. *Development* **118**:1209–1219.
- NEI, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.
- NEI, M., and T. GOJOBORI. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**:418–426.
- ORTEGA, M.-A., M. DÍAZ-GUERRA, and L. SASTRE. 1996. Actin gene structure in two *Artemia* species, *A. franciscana* and *A. parthenogenetica*. *J. Mol. Evol.* **43**:224–235.
- OSUNA, C., and J. SEBASTIÁN. 1980. Levels of the RNA polymerase during early larval development of *Artemia*. *Eur. J. Biochem.* **109**:383–389.
- PETERSON, G. L., R. D. EWING, S. R. HOOTMAN, and F. P. CONTE. 1978. Large scale partial purification and molecular and kinetic properties of the (Na+K)-activated adenosine triphosphatase from *Artemia salina* nauplii. *J. Biol. Chem.* **253**:4762–4770.
- ROZAS, J., and R. ROZAS. 1997. DnaSP version 2.0: a novel software package for extensive molecular population genetics analysis. *Comput. Appl. Biosci.* **13**:307–311.
- SANGER, F., S. NICKLEN, and A. R. COULSON. 1977. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463–5467.

- SASTRE, L. 1999. Isolation and characterization of the gene coding for *Artemia franciscana* TATA-binding protein: expression in cryptobiotic and developing embryos. *Biochim. Biophys. Acta* **1445**:271–282.
- SERRANO, R. 1989. Structure and function of plasma membrane ATPase. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **40**:61–94.
- SHETERLINE, P., and J. C. SPARROW. 1994. Protein profile: actin. Academic Press, London.
- SIMONSEN, K. L., G. A. CHURCHILL, and C. F. AQUADRO. 1995. Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* **141**:413–429.
- SOKAL, R. R., and F. J. ROHLF. 1995. *Biometry*. Freeman, San Francisco.
- STROBECK, C. 1983. Expected linkage disequilibrium for a neutral locus linked to a chromosomal arrangement. *Genetics* **103**:545–555.
- SUN, D. Y., J. Z. GUO, H. A. HARTMANN, H. UNO, and L. E. HOKIN. 1992. Differential expression of the Alpha2 and Beta messenger RNAs of the Na,K-ATPase in developing brine shrimps as measured by in situ hybridization. *J. Histochem. Cytochem.* **40**:555–562.
- TAJIMA, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**:585–595.
- VASILETS, L. A., and W. SCHWARTZ. 1993. Structure-function relationships of cation binding in the Na/K-ATPase. *Biochim. Biophys. Acta* **1154**:201–222.

CARO-BETH STEWART, reviewing editor

Accepted October 11, 1999

APPENDIX

Alignment of a Fragment from the $\alpha 1$ Na/K-ATPase Coding Region from *Artemia franciscana* (c136) and *Artemia parthenogenetica* (Apa)

c136	ATGGCAAAG	GCAAGCAAAA	GAAAGGAAAG	GATTTGAATG	AATTGAAGAA	GGAGTTGGAT	ATTGACTTTC	ATAAAATTCC	148
ApaT..C.C..	
								Ile	
c136	TATTGAAGAA	TGTTATCAAA	GGCTTGAAG	CAATCCTGAA	ACCGGTTTAA	CTAATGCTCA	AGCTCGTAGT	AATATTGAAC	228
ApaG....	
								Met	
c136	GTGATGGTCC	CAACTGTTTA	ACCCACCTA	AAACAACCC	TGAATGGATC	AAGTTTTCGA	AGAACCTTTT	TGGTGGTTTT	308
Apa	G.....C..A..T....	
c136	GCATTACTCC	TTGGACAGG	CGCTATTTTA	TGCTTCTTG	CCTATGGAAT	TGAAGCTTCA	TCAGGAAATG	AAGATATGCT	388
ApaG	..T..C....C.....	..G..C.....	
c136	GAAGATAAC	CTTACCTTG	GAATAGTGT	GGCAACTGT	GTCATTGTGA	CTGGCATCTT	TTCGTATTAT	CAAGAAAACA	468
Apa	A.....T....T....	
				Leu					
c136	AATCCTCCCG	TATTATGGAC	AGCTTCAAGA	ACTTGGTCCC	CCAG	512		GTA	768
ApaTA....G..	
				Met					
c136	CAATGAGAG	TATTGTGATT	GGTATGGTG	ACAACACTGT	TATGGGTCGT	ATTGCTGGTC	TCGCCTCTGG	ACTGGACACT	848
Apa	G.....G...	
								Glu	
	Glu								
c136	GGAGAGACGC	CTATTGCCAA	AGAAATCGCC	CACTTTATTC	ACATTATTAC	GGGTGTCGCT	GTGTTTCTTG	GTGTCACTTT	928
ApaC....T.....	T..C.....	
	Asp								
c136	CTTCATTATT	GCTTTTGTCC	TTGGATATCA	CTGGCTGGAT	GCTGTAGTCT	TTCTAATTGG	TATTATCGTC	GCTAATGTTC	1008
ApaAA.	
								Lys	
c136	CTGAAGGTCT	TCTAGCCACT	GTCACTGTTT	GTTTGACCCT	GACAGCCAAA	CGTATGGCTT	CCAAAACTG	TCTCGTCAAA	1088
Apa	.C.....A.	A.....T..T....	
								Asn	
c136	AATCTCGAAG	CCGTCGAAAC	CCTTGGCTCC	ACCTCAACAA	TTTGCTCAGA	TAAGACTGGC	ACCCTGACTC	AAAACCGTAT	1168
ApaT....C...C..A.	
c136	GACAGTCGCT	CACATGTGGT	TCGACGGAAC	GATCACCGAG	GCTGATACTA	CAGAAGATCA	GTCAGGGGCT	CAGTTTGATA	1248
ApaT....T...T.....	
c136	AATCATCTGC	TGGATGGAAA	GCTCTTGTTA	AAATTGCCGC	CCTTTGTAGT	CGTGCTGAAT	TCAAGCCTAA	TCAAAGCACC	1328
ApaG..G.	T.....C...NNNNNNNN	NNNNNNNNNN	
c136	ACCCCAATTT	TGAAAAGGGA	AGTAACTGGT	GATGCCTCTG	AAGCTGCTAT	TTTGAAATGC	GTGGAATAA	CCACTGGCGA	1408
Apa	NNNNNNNNNN	NNNNNNNNC...TG.	
c136	AACTGAAGCT	ATTCGAAAGA	GGAACAAGAA	GATCTGTGAA	ATCCCATFCA	ACTCTGCCAA	CAAATTTTCA	1478	
ApaC.....	.A.....	A..A.....C....	.T.....	T.....		

NOTE.—Several fragments of cDNA 136 are shown here aligned to the homologous partial sequences from *A. parthenogenetica* obtained. The starting methionine codon is in bold letters. The replacement substitutions are indicated with amino acids situated by the corresponding codon position. The numbers on the right refer to the complete sequence of cDNA 136 (accession number X56650).