Copyright © 2001 deNovo Biologic LLC. All rights reserved.

An oxygen-related bioprocess drives eukaryote-to-prokaryote genome evolution and speciation

Douglas H. Robinson

deNovo Biologic LLC
5528 North 24th Street
Arlington, Virginia 22205-3313 USA

Prokaryotes were isolated from a sterile cell culture system in which sterile retrovirally transformed human eukaryotes were subjected to an oxygen environment of alternating anaerobiosis and aerobiosis. All prokaryotes were isolated within 15 days of bioprocess initiation and all were classified as Gram-positive eubacteria, bacilli and staphylococci, which were often highly pleomorphic in culture. This sterile bioprocess was validated in an independent study. Genome sizing by pulsed-field gel electrophoresis indicated that the genomes of representative prokaryotes have approximate sizes of 1.6 megabases and 2.1 megabases. Through DNA sequencing and *in silico* analysis, one representative prokaryotic genome was found to contain transposable and genetic elements that match DNA sequences which map to 19 different human chromosomes (1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 15, 17, 18, 19, 20, 21 and X), to staphylococcal genomes and to plant genomes. Many representative prokaryote genomic DNA sequences analyzed were found to be novel with no significant matches to DNA sequences contained in current nucleotide sequence databases. As the human genome nucleotide sequence database expands and/or the methods of *in silico* analysis advance, some novel DNA sequences, many with apparent coding regions, might be found to match DNA sequences that map to the human genome. Other novel DNA sequences might be the result of widespread recombination events involving human retroelements. A few representative prokaryote genomic DNA sequences match DNA sequences that map to both the staphylococcus genome and the human genome, indicating that eubacterial DNA elements are present in the human genome. In addition, immunoblot studies indicated that various prokaryotes so isolated express both human genes/proteins that map to many different human chromosomes and retroviral proteins. Prokaryotes were isolated from other human and mammalian eukaryotic cell lines with this culture system. Biocomputing analogies can be used to explain the associated eukaryote-to-prokaryote genome evolution which appears to be speciation-related and retrolement-related. Although questions remain to be answered regarding the mechanism(s), the simplicity of this oxygen-related bioprocess, a type of directed reductive evolution, lends readily to the development of applications including the rapid generation of bacterial expression systems, human and animal biologicals, novel chimerical proteins and vaccines. [The sequence data described in this paper have been submitted to the GenBank data library under accession numbers AF266540-AF266551 and AF286220-AF286221.]

eukaryote | prokaryote | Gram-positive eubacteria | human chromosomes | human genome | oxygen | bioprocess | retroelements | SINEs | LINEs | retrovirus | DNA transfer | genome evolution | cellular evolution | directed evolution | speciation | tumorigenesis | DNA sequencing | *in silico*

Introduction

A popular theory of cellular evolution suggests that many modern-day eukaryotes are the products of ancient endosymbioses involving free-living prokaryotes[1-3]. Phylogenetically, Gupta and Doolittle have considered the eukaryotic nuclear genome to be an evolutionary chimera, containing both eubacterium-related genes and archaebacterium-related genes[4-6]. Based upon new approaches to molecular phylogenetic and biochemical reconstructions of cellular evolution and contrary to endosymbiotic scenarios for eukaryotic genome evolution, Forterre, Poole and others have proposed an ancient eukaryotic rooting for the universal tree of life in which the prokaryotic genome is derived from a eukaryote-like genome[7-13].

Assuming that the eukaryotic genome contains prokaryotic 'information' (through endosymbiosis or not) and that changes in environment increase the rate of biologic evolution[14], I hypothesized that changes in environment might induce the orthologous, 'vertical' genomic DNA transfer associated with a modern-day eukaryote-to-prokaryote genome evolution *and* speciation, a higher order process that appears to be inextricable from eukaryote-to-prokaryote genome evolution. In addition, I assumed that the bioprocess might occur saltationally, *i.e.* within a *rapid and observable* timeframe. The latter assumption counters the common assumption in cellular evolutionary theory that all extant life including all prokaryotic lineages have evolved continuously from an ill-defined 'last universal common ancestor' ('LUCA') that existed in an unobservable distant past[8,9].

Fruitful direction was found in consideration of the critical role of retroviral elements in environmental stress-induced macroevolutionary events[15]. As an environmental pressure, anoxia is a potent stimulus for the induction of various retroelements[16,17]. Correspondingly, changes in planetary oxygen atmosphere may have been associated with symbiogenesis and the attendant genomic DNA transfer in early cellular evolution[1]. After considerable empirical testing, I developed a sterile bioprocess using a sterile *in vitro* cell culture system in which sterile retrovirally transformed human eukaryotes were subjected to an environmental pressure of alternating aerobiosis and anaerobiosis. This empirical, experimental method could be viewed as a test of a recent theoretical prediction that a prokaryotic genome can be derived from a eukaryote-like genome[8] with the additional premise that this genome evolution is inextricable from cellular evolution or speciation.

Within 15 days of bioprocess initiation, pleomorphic prokaryotes with Gram-positive eubacterial features were isolated from sterile retrovirally transformed human eukaryotes using this sterile bioprocess. The following results and discussion report the findings and arguments supporting the conclusion that this sterile bioprocess drives eukaryote-to-prokaryote genome evolution and speciation.

© 2001 deNovo Biologic | www.denovo-bio.com

Results

Sterility Testing of the Retrovirally Transformed Human Eukaryotic Cell Line, ATCC CRL 11655, Indicates that the Cell Line Is Free of Microbial Contamination

Extensive in-house and independent sterility testing by BioReliance Corp. (Rockville, MD) indicated that the ATCC CRL 11655 cells (Fig. 1a) used in these experiments harbored no microorganisms prior to exposure to the oxygen-related environmental pressure. Sterility studies included a highly sensitive, double-step-PCR molecular phylogenetic method using nested, degenerate primers and positive controls to detect gene sequences coding for the evolutionary conserved 16 S rRNA of some twenty-five mycoplasmas, including those most commonly found in cell cultures[18].

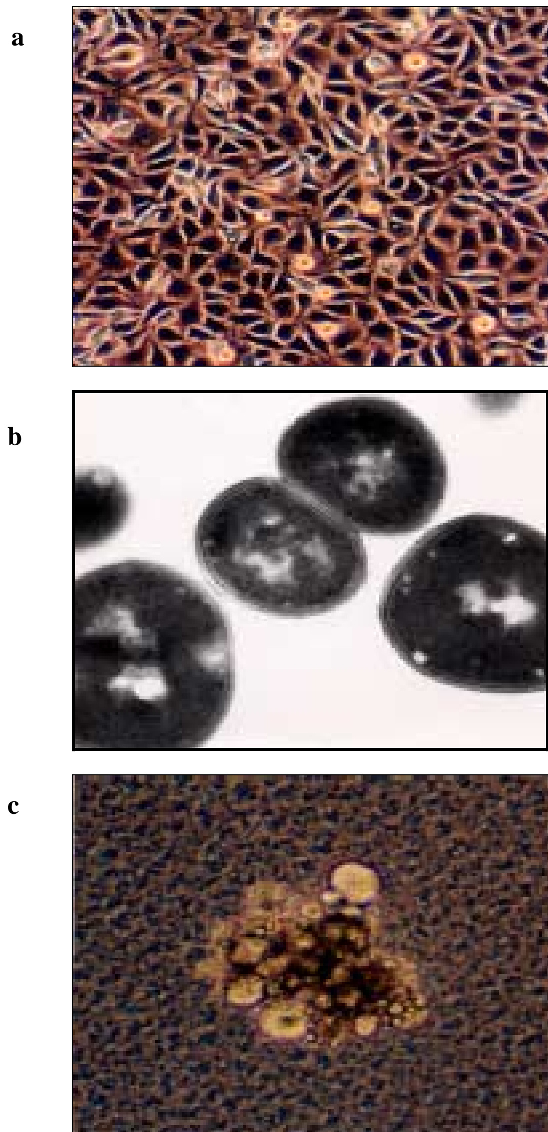


Figure 1 Photomicroscopies of retrovirally transformed human eukaryotic cells and a prokaryote isolated from the cell culture system. **a**, light photomicroscopy of the retrovirally transformed human eukaryotic cell line in culture actively propagating under standard aerobic cell culture conditions (magnification, x 25), **b**, transmission electron photomicroscopy of a prokaryote isolated from the cell culture system (magnification, x 45,900), **c,** fungus-like morphologies observed for some prokaryotes (magnification, x 37.5).

Prokaryotes Were Isolated from the Sterile Cell Culture System Within 15 Days of Bioprocess Initiation

The sterile cell culture system involved exposing sterilely vented culture flasks of sterilely cultured eukaryotic cells to sterile alternating aerobic/anaerobic conditions (the environmental pressure phase) followed by sterile aerobic bacteriological cell culture of flask contents (the propagation phase). Approximately one prokaryotic colony was isolated during the propagation phase for each 7.5 x 10^7 ATCC CRL 11165 cells exposed to the oxygen-related environmental pressure in the environmental pressure phase. All colonies were isolated within 15 days of the initiation of experiments. *Prokaryotes were isolated from no other condition* including medium-only controls in the modified anaerobic jar and controls in the environmental pressure phase in which eukaryotic cells were incubated in a stable atmosphere of air/5% CO_2 (see Methods). Prokaryotes were isolated from this sterile cell culture system in an extensively-controlled independent study by BioReliance Corp. (see Methods). A total of 19 different prokaryotic colonies were isolated in 11 experiments using various mammalian eukaryotic cell lines including those human, murine , porcine and derived from tumors.

Prokaryotes Isolated with the Bioprocess Have Gram-Positive Eubacterial Features and Are Often Pleomorphic

All prokaryotes were classified *initially* as various species of Gram-positive eubacteria. Morphologies observed include those coccal (Fig. 1b), bacillary and, often, highly pleomorphic and fungus-like (Fig. 1c). Prokaryote W, an isolate selected for in-depth molecular analysis, was identified by biochemical typing at the ATCC, *with difficulty and tentatively*, as a strain of *Staphylococcus capitis. Staphylococcus capitis* is a member of the *Staphylococcus epidermidis* species group[19].

Genome Sizes of Pleomorphic Prokaryote W Colony Variants Differ by Several Kilobases

The prokaryote P genome size was determined to be approximately 2.1 megabases by one-dimensional pulsed-field gel electrophoresis (not shown)[20]. Prokaryote W was highly pleomorphic, exhibiting two colony variants during plate culture for pulsed-field gel electrophoresis. This colony-variant pleomorphism in plate culture was observed with other prokaryotes derived with this bioprocess. Pulsed-field gel electrophoresis indicated that the prokaryote W colony variants have genome sizes of approximately 1.6 megabases and 2.1 megabases.

The Riboprint of One Prokaryote W Colony Variant Does Not Match Any Known Genome Size of Its Species Group

Riboprinting or automated ribotyping, the automated determination of DNA sequence restriction patterns for prokaryotic rRNA, can be useful in the classification of prokaryotes into species groups. Riboprinting for the prokaryote W colony variant with a genome size of approximately 2.1 megabases was indeterminate with respect to existing riboprint databases. The riboprint for the prokaryote W colony variant with a genome size of approximately 1.6 megabases indicated that this colony variant belongs to the *Staphylococcus epidermidis* species group (not shown). However, no average genome size for this staphylococcus species group has been observed to be less

© 2001 deNovo Biologic | www.denovo-bio.com

than approximately 2.2 megabases[19]. This mismatching of riboprint and prokaryote W genome size with reference to existing literature and the difficulties in riboprinting in general highlight the novelty of the often pleomorphic prokaryotes derived with this bioprocess. Prokaryote P riboprinting was also indeterminate.

Immunoblotting Studies Indicate that Prokaryotes Isolated Express Murine Retroviral and 'Human' Proteins

Prokaryotes derived with this bioprocess were examined for the presence of murine retroviral genes/proteins from the murine retrovirus used to establish the retrovirally transformed human eukaryotic cell line. Prokaryotes and retrovirally transformed human eukaryotic cells express several retroviral *gag* precursor polyproteins in addition to the retroviral gag p30 core protein[21,22] (Fig. 2a). This finding was confirmed by restriction mapping of a 500-bp murine retroviral *gag* gene product, PCR-amplified from retrovirally transformed human eukaryotic cell DNA templates and from prokaryotic DNA templates (not shown). One prokaryote expresses *and secretes* two isoforms of human serum albumin (Fig. 2b). Immunoblot studies indicate that other prokaryotes isolated express a variety of human proteins (not shown).

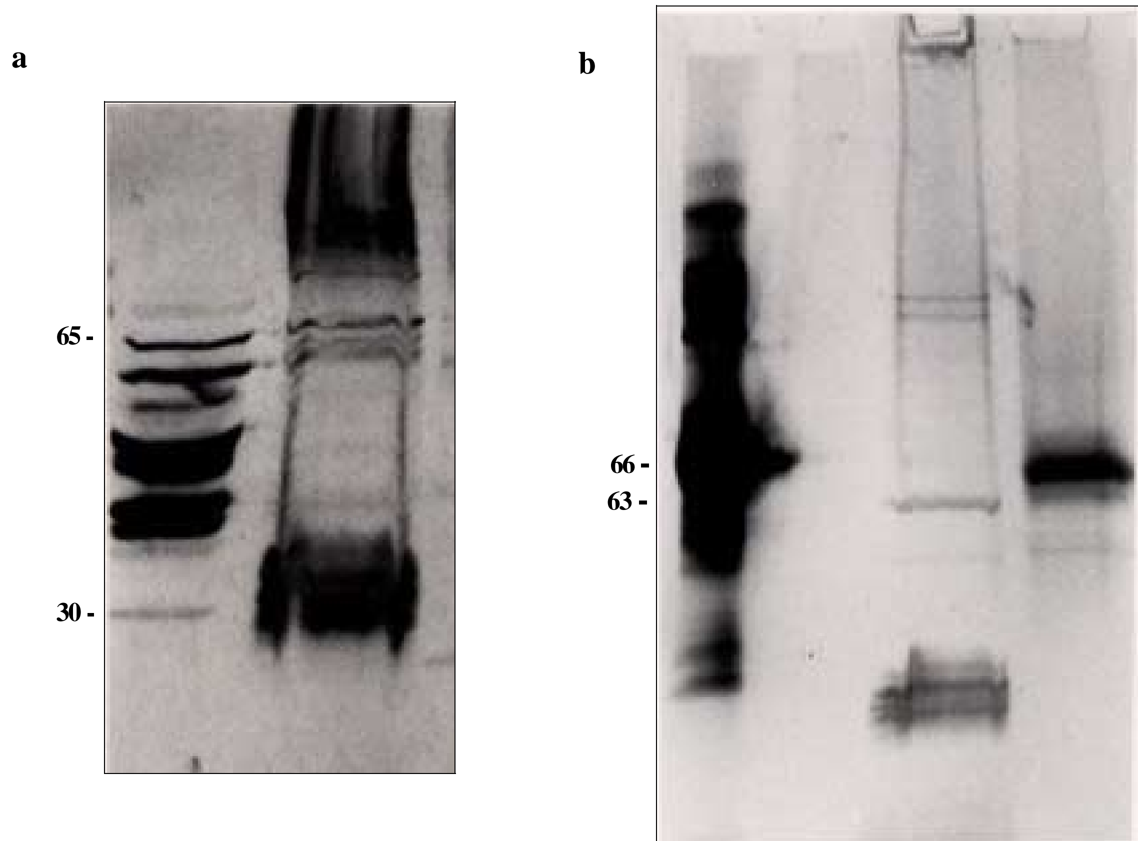
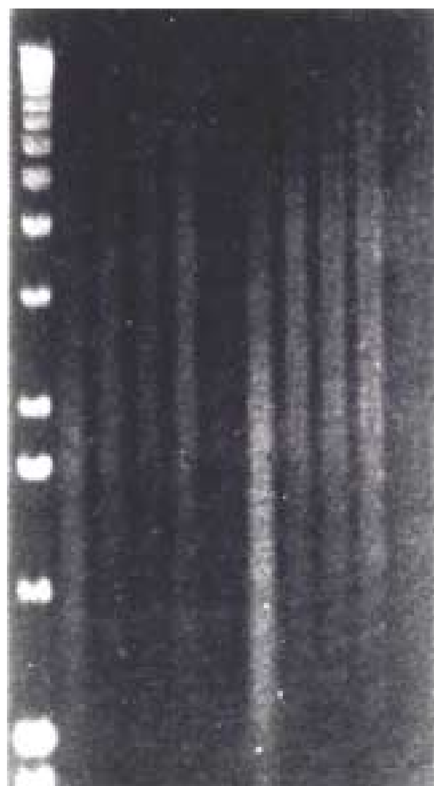


Figure 2 Detection of the L-cell virus p30 core protein and related polyproteins. **a,** Immunoblot for murine retroviral *gag* p30 core protein. Lanes: 1, retrovirally transformed human eukaryotic cell protein pellet; 2, a protein pellet from a prokaryote isolate used for the *gag* gene PCR and restriction enzyme analysis; **b,** Immunoblot of human serum albumin (HSA). Lanes: 1, HSA standard (1.5 fg); 2, open; 3, prokaryotic protein pellet; 4, prokaryotic culture supernatant protein pellet. This blot was probed with a monoclonal, anti-human serum albumin antibody that recognizes 63 and 66 kD HSAs. Note that the prokaryotic protein pellet sample contains the 63 kD isoform and related polyproteins. The supernatant protein pellet sample contains the 63 kD isoform and the 66 kD isoform.

© 2001 deNovo Biologic | www.denovo-bio.com

Human Alu/LINE-1-directed PCR Amplification of Prokaryotic Genomic DNA Suggests that Human Genome-Related Transposable Elements Are Present

Human *Alu*/LINE-1-directed PCR amplification of prokaryotic genomic DNA was used to screen for the presence of human genome-related transposable/repeat elements. Gel electrophoresis DNA smears typical of human *Alu*/LINE-1-directed PCR amplification were observed when using 'human-specific' *Alu* and LINE-1 primers[23, 24] with genomic DNA templates from human placenta, the retrovirally transformed human eukaryotic cell line, and two prokaryotes, P (ATCC 55589) and W (ATCC 55590) (Fig. 3). No *Alu* and LINE-1 inter-repeat element DNA smears were observed when using control porcine genomic DNA templates. The presence of these DNA smears suggests that many sequences are amplified, obscuring gel-electrophoresis visualization of individual sequences[23].



1 2 3 4 5 6 7 8 9 10 11

Figure 3 Gel electrophoresis DNA smears typical of human *Alu*/LINE-1-directed PCR of inter-repeat elements. Lanes: 1, 1 kb ladder (BRL); 2, human placenta; 3, retrovirally transformed human endothelial cell; 4, prokaryote P; 5, prokaryote W; 6, porcine eukaryotic cell control; 7, human placenta; 8, retrovirally transformed human endothelial cell; 9, prokaryote P; 10, prokaryote W; 11, porcine eukaryotic cell control. *Alu* and human LINE-1 inter-repeat element DNA smears in lanes 2-6 were amplified using oligonucleotide primers designated '517' [23] and 'L1S' [24]. Smears in lanes 7-11 were amplified using oligonucleotide primers designated 'TC-65'[23] and, again, 'L1S.' The molecular weights of the *Alu*/LINE-1 PCR inter-repeat element DNA products were in the kilobase range.

DNA Sequencing and In Silico Analysis of Cloned Human Alu LINE-1-directed PCR Amplification of DNA Products from Prokaryotic Genomic DNA Reveals Human Genome-Related DNA

DNA products from a human *Alu*/LINE-1-directed PCR amplification of prokaryotes P and W genomic DNA using oligonucleotide primers 'TC65' and 'L1S'[23, 24] were cloned with a T/A cloning method and sequenced by primer extension. Fifty PCR DNA products from each amplification were sequenced. Using the National Center for Biotechnology Information (NCBI) BLASTN and BLASTX programs[25] with the GenBank nonredundant (nr) nucleotide sequence database, a number of the PCR DNA sequences from both the genomes of prokaryotes P and W matched DNA sequences that map to the human genome. These PCR DNA sequences include both human genome-related extragenic repeat elements (SINEs, LINEs, MIRs, *etc.*) and human genome-related non-repeat genetic elements. The prokaryote W genome was selected for in-depth *in silico* analysis.

DNA Sequencing and In Silico Analysis of Cloned Human Alu/LINE-1-directed PCR Amplification of DNA Products from Prokaryote W Genomic DNA Reveal 22 Kilobases of DNA Related to 19 Different Human Chromosomes

Human *Alu*/LINE-1-directed PCR DNA sequences amplified from the prokaryote W genome were found to match DNA sequences that map to specific human chromosomes. Fig. 4 and 5a-c depict PCR DNA sequence matches that map to human chromosomes 9, 11, 12, 20, 21 and X and a human kpni repeat mRNA cDNA sequence using the BLASTN and BLASTX programs with the GenBank nonredundant (nr) nucleotide database (see Appendix I for a representative sequence alignment). Matches to DNA sequences from chromosomes 9, 11, 12 and X and the kpni repeat mRNA cDNA were found to have Expect values (E values) of 0.0 . The E value is a parameter or statistical significance threshold that describes the number of hits (or "background noise") one can expect to see just by chance when searching a database of a particular size for sequence matches[25]. For the database searches described, the default value of 10 was used, such that 10 matches are expected to be found merely by chance.

Using the BLASTN program with the GenBank high throughput genome sequencing (htgs) nucleotide sequence database, other *Alu*/LINE-1-directed PCR DNA sequences amplified from the prokaryote W genome were found to match DNA sequences that map to human chromosomes 2, 4, 5, 6, 9, 10, 11, 12, 17 and 21 (E values: 0.0) and human chromosomes 1 (E value: 1e-102), 3 (E value: 2e-141), 8 (E value: 1e-99), 13 (E value: 8e-82), 15 (E value: 2e-91), 18 (E value: 2e-100) and 19 (E value: 1e-122). Most of these htgs-matched PCR DNA sequences were found to contain nonrepeat genetic elements when using repeat masking to exclude repeat DNA sequences from the target DNA sequences during *in silico* analysis. Altogether, 22 kilobases of human *Alu*/LINE-1-directed PCR DNA sequences from the prokaryote W genome were found to be related to DNA sequences that map to these 19 different human chromosomes (1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 15, 17, 18, 19, 20, 21 and X; E values of 0.0 for sixteen DNA sequence matches).

The prokaryote P genome was found to contain human *Alu*/LINE-1-directed PCR DNA sequences that match DNA sequences that map to at least 10 different human chromosomes (not shown). Other *Alu*/LINE-1-directed PCR DNA sequences from the prokaryote P genome appeared to be human but indeterminate as to specific human chromosomal origin.

© 2001 deNovo Biologic | www.denovo-bio.com

A Few Human Alu/LINE-1-directed PCR Amplification DNA Products from Prokaryote W Genomic DNA Are Prokaryote-Related DNA Sequences

Several kilobases of human *Alu*/LINE-1-directed PCR DNA sequences were found to match DNA sequences that map to the *Staphylococcus epidermidis* genome through BLASTN analysis (as in Fig. 5d). Several PCR-product DNA sequences did not have matches to human or *Staphylococcus epidermidis* genomic DNA sequences using current nucleotide databases. Some prokaryote W nucleotide sequences match DNA sequences of proteins involved in *Staphylococcus aureus* cell wall synthesis (Fig. 5e). Sequence identities of these prokaryote W nucleotide sequences to these known DNA sequences are moderate with matches varying from 71% to 82%. It is possible that these *Staphylococcus aureus-like* sequences will be found to match *Staphylococcus epidermidis* genome sequences in future studies. Nonetheless, these PCR DNA sequence amplification data indicate that the prokaryote W genome is chimerical, consisting of human-like DNA sequences admixed with prokaryote-like DNA sequences. This observation may have some relationship to protein phylogenies that indicate a close evolutionary relationship between Gram-positive eubacteria and the probable existence of 'chimerical' prokaryotic genomes in nature [4, 5, 26].

DNA Sequencing and In Silico Analysis of a Shotgun Library Sampling of the Prokaryote W Genome Indicates That 60% of Prokaryote W Shotgun Library DNA Sequences (67 kilobases of DNA) Are Novel

Next, a shotgun library of prokaryote isolate W genomic DNA was constructed to obtain a random sampling of genomic DNA fragments for sequencing and *in silico* analysis. 109 kilobases of DNA were sequenced. 60% of the DNA sequences from the shotgun library sampling (67 kilobases of DNA) have no significant matches to DNA sequences (E values < 1e-20) contained in the current NCBI nucleotide sequence databases using the NCBI BLASTN program.

Other Shotgun Library Prokaryote W DNA Sequences Are Related to Prokaryote Genetic Elements, to Both Prokaryotic and Human Genetic Elements and to Plant Genetic Elements

Approximately 37% of the DNA sequences from the shotgun library sampling (40 kilobases of DNA) match DNA sequences that map to the *Staphylococcus epidermidis* genome (E values: 0.0) using the NCBI BLASTN program. Prokaryote W was typed as a strain of *Staphylococcus capitis*, which belongs to the *Staphylococcus epidermidis* species group. Figure 6a depicts one such prokaryote W shotgun library DNA sequence that matches a Staphylococcus *epidermidis* DNA sequence that contains 16S and 23S rRNA DNA sequences. This finding agrees with one prokaryote W riboprinting (see above). Multiple regions of this DNA sequence match htgs-database DNA sequences that map to human chromosomes 5 and 16 DNA sequences submitted by the Production Sequencing Facility of the DOE Joint Genome Institute.[27] The *in silico* analysis of these *informational* gene-related DNA sequences suggests that some nucleotide sequences within the human genome may be "eubacterial" in nature. Two prokaryote W shotgun library DNA sequences are shown that match DNA sequences that map to the *Staphylococcus aureus* SRImec-I and SCCmec-I region (Fig. 6b) and a DNA sequence of the *Staphylococcus epidermidis* genome (Fig. 6c). These DNA sequences are representative of prokaryote W DNA sequences that appear to be staphylococcal in nature.

One novel shotgun library prokaryote W genomic sequence matches a DNA sequence that maps to an mRNA of a putative Hs1pro-1 homologue from *Glycine max* (soybean) using analysis with the GenBank nr nucleotide sequence database (Fig. 6d). Another novel shotgun library prokaryote W genomic sequence matches a DNA sequence that maps to a cytochrome p450 mRNA from *Vigna radiata* (mung bean) (not shown). Recently, a human gene similar to plant lysophosphatidic acid acyltransferase was discovered within human chromosome 21[28]. These findings suggest that the prokaryote W genome contains homologues of plant genes that were transferred from the human genome. Since the Hs1pro-1 homologue corresponds to an mRNA sequence, its transfer may have involved reverse transcription. This mRNA sequence corresponds to less than half of an approximate three kilobase DNA sequence. The remainder of this DNA sequence is novel with no significant matches to DNA sequences contained in current NCBI nucleotide sequence databases (Fig. 6d, second figure). Other prokaryote W shotgun library DNA sequences were found to match htgs-database DNA sequences that map to human chromosomes 9 and 17 (not shown).

Discussion

The Evidence Indicates a Process of a Eukaryote-to-Prokaryote Genome Evolution That Involves a Vertical Genomic DNA Transfer of Speciation Rather Than Eukaryote-to-Prokaryote Horizontal Genomic DNA Transfer

Little, if any, literature documents experimental environmental pressure-induced eukaryote-to-prokaryote horizontal DNA transfer. Most studies of eukaryote-to-prokaryote horizontal DNA transfer involve historical, non-empirical molecular phylogenetic reconstructions of discrete genomic DNA transfers[6]. A recent review of horizontal (lateral) gene transfer in bacteria emphasizes prokaryote-to-prokaryote horizontal gene transfer and notes the importance of transposon-related sequences in horizontal gene transfer in bacteria[29]. However, few specifics are given regarding eukaryote-to-prokaryote horizontal gene transfer and no mention is made of experimental environmental pressure-induced eukaryote-to-prokaryote horizontal DNA transfer.

The *in vitro* and *in silico* evidence presented documents a representative prokaryotic genome that contains *at least* 22 kilobases of genomic DNA sequences, both genetic and extragenic, which match DNA sequences that have been mapped to 19 different human chromosomes (1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 15, 17, 18, 19, 20, 21 and X; E values of 0.0 for sixteen DNA sequence matches). This finding indicates widespread inter-chromosomal recombination of human genome-related DNA as part of the genomic DNA transfer process. The human *Alu*/LINE-1-directed PCR DNA smear and DNA sequence data indicate that numerous human genome-related retroelements are involved in the genomic DNA transfer and that prokaryote-related genetic elements are admixed with human genome-related extragenic and genetic elements in these prokaryotic genomes. These findings suggest that the genomic DNA transfer may be retroelement-mediated.

In addition to the murine retrovirus present in this human cell line, *Alu* and LINE-1 retroelements (interspersed repeats) are abundantly available in the human genome to serve as a retroelement-based infrastructure for a eukaryote-to-prokaryote genome reorganization. Recent sequencing of human chromosomes 21 and 22 reveals that these retroelements together make up approximately one-fourth

© 2001 deNovo Biologic | www.denovo-bio.com

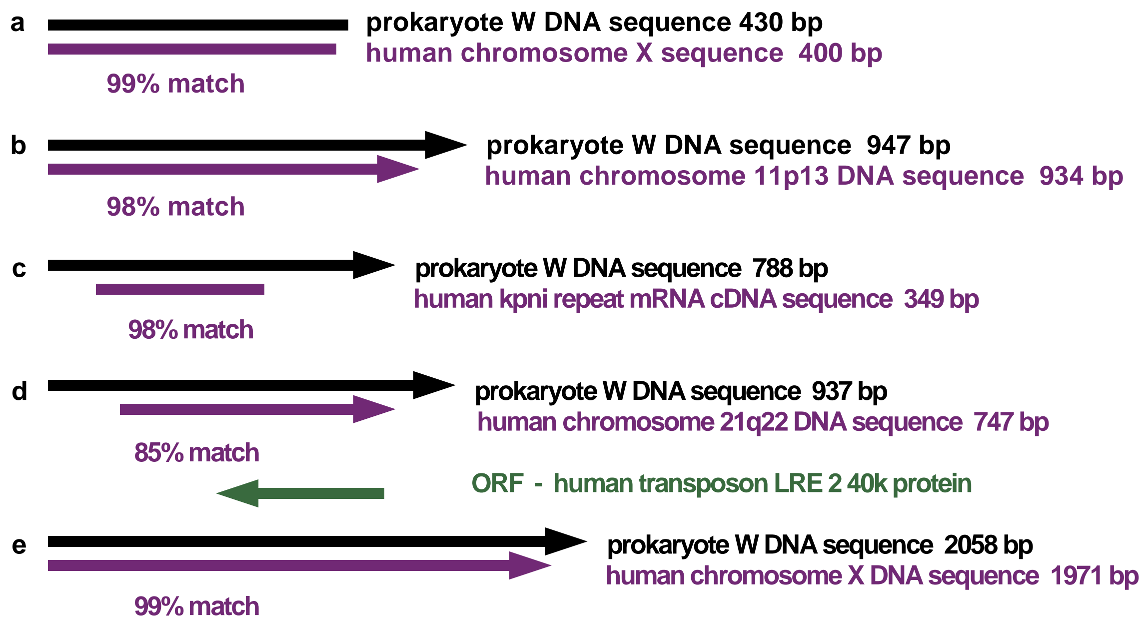


Figure 4 Sequence analysis of inter-repeat element DNA products from a human *Alu*/LINE-1-directed PCR amplification of prokaryote W genomic DNA using oligonucleotide primers "TC65" and L1S." Matches (or *percent* identities) map to DNA sequences contained in the GenBank nonredundant (nr) database. **a,** a prokaryote W DNA sequence (GenBank accession number AF266540) that matches a DNA sequence that maps to human chromosome X and contains an Alu-Sb repeat subfamily and other human repeat sequences (E value: 0.0) (GenBank accession number AC004074), **b**, a prokaryote W DNA sequence (GenBank accession number AF266541) that matches a DNA sequence that maps to human chromosome 11p13 (E value: 0.0) and contains LINE-1 and Alu-Sx subfamily repeats (EMBL accession number HSDJ259N9), **c,** a prokaryote W DNA sequence (GenBank accession number AF266542) that matches a cDNA sequence that maps to a human kpni (LINE-1) repeat mRNA (E value: 0.0) (GenBank accession number HUMRSKP08), **d,** a prokaryote W DNA sequence (GenBank accession number AF266543) that matches a DNA sequence that maps to human chromosome 21q22 and contains a LINE-1 repeat encoding a peptide with homology (42%-44% si by BLASTX analysis) to a number of human p40 proteins (highest E values: 9e-172 for the nucleotide sequence / 3e-44 for the protein sequence) (GenBank accession number AF020803; PIR accession number I38587) and **e,** a prokaryote W DNA sequence (GenBank accession number AF266544) that matches a DNA sequence that maps to human chromosome X (E value: 0.0) (EMBL accession number HSU227D1). See Appendix I for a representative sequence alignment. In some sequence series, short duplicated repeats flank the sequence with an E value of 0.0. Sequences upstream and downstream to these repeats have lower E values.

© 2001 deNovo Biologic | www.denovo-bio.com

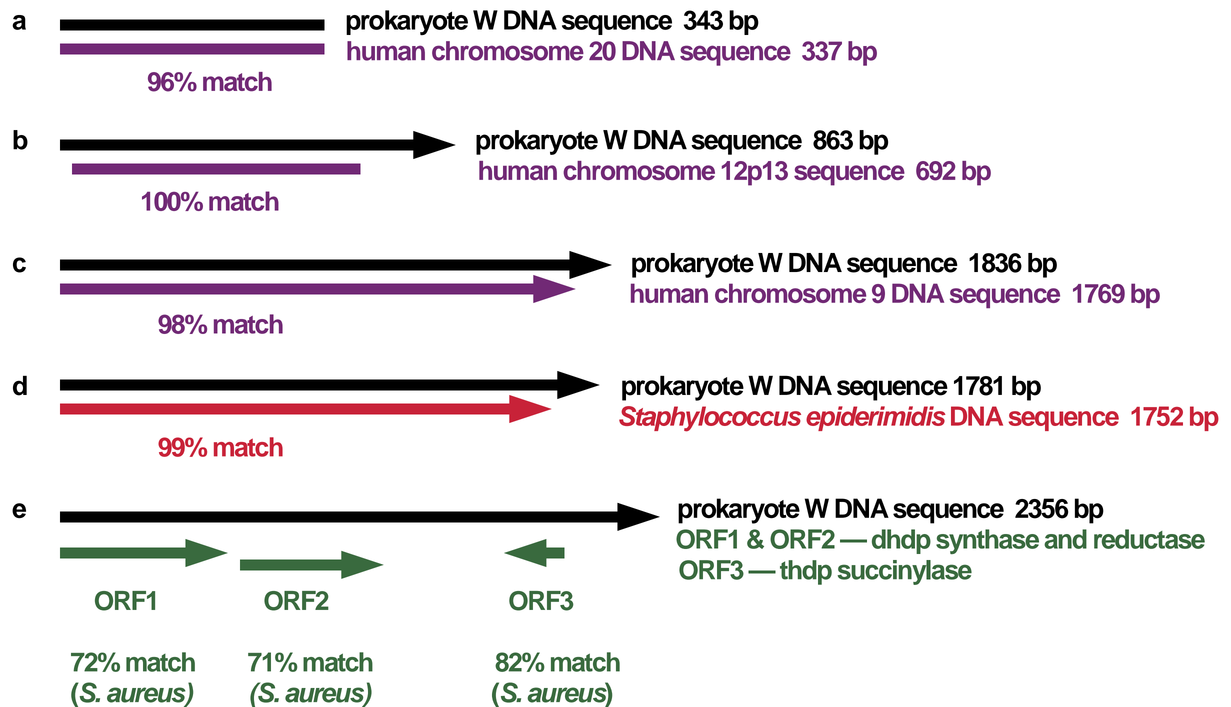


Figure 5 Sequence analysis of inter-repeat element DNA products from a human *Alu*/LINE-1-directed PCR amplification of prokaryote W genomic DNA using oligonucleotide primers "TC65" and L1S." Matches (or *percent* identities) map to DNA sequences contained in the GenBank nonredundant (nr) database. **a**, a prokaryote W DNA sequence (GenBank accession number AF266545) that matches a DNA sequence (L1MC2 repeat) that maps to human chromosome 20 (highest E value: 2e-83) (EMBL accession number HSJ944N18), **b,** a prokaryote W DNA sequence (GenBank accession number AF266546) that matches a DNA sequence that maps to human chromosome 12 (E value: 0.0) (GenBank accession number AC012156) and contains MIR and MSTA repeats, **c,** a prokaryote W DNA sequence (GenBank accession number AF266547) that matches a DNA sequence that maps to human chromosome 9 (E value: 0.0) and contains the repeat families AluY, L1P1, MIR, L1M4 and L1PA6 in addition to non-repeat sequence (GenBank accession number AC006313), **d,** a prokaryote W DNA sequence (GenBank accession number AF266549) that matches a DNA sequence that maps to the *Staphylococcus epidermidis* genome (E value: 0.0) (GenBank accession number AF269663) and **e,** a prokaryote W DNA sequence (GenBank accession number AF266548) that matches a DNA sequence that maps to a DNA sequence that contains a tandem repetition of genes/proteins homologous to dihydrodipicolinate (dhdp) synthase and reductase and tetrahydrodipicolinate (thdp) succinylase in *Staphylococcus aureus* (E values: ORF1 - e-120, ORF2 - 3e-88 and ORF3 - 2e-83) (ORFs 1, 2 and 3 - GenBank accession number AF306669). In some sequence series, short duplicated repeats flank the sequence with an E value of 0.0. Sequences upstream and downstream to these repeats have lower E values.

© 2001 deNovo Biologic | www.denovo-bio.com

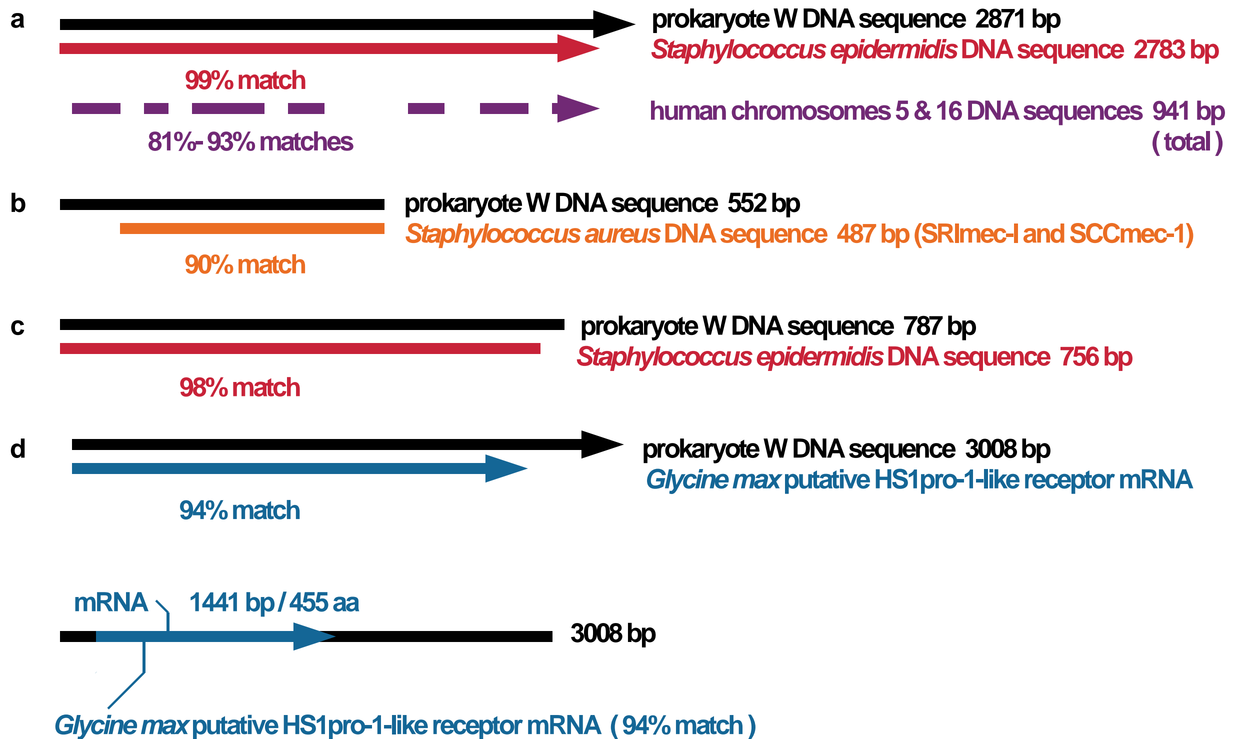


Figure 6 Shotgun library sequences from the prokaryote W genome. Matches (or *percent* identities) map to DNA sequences contained in the GenBank nonredundant (nr) database and the GenBank htgs database. **a,** a prokaryote W DNA sequence (GenBank accession number AF266550) matches a DNA sequence that maps to the *Staphylococcus epidermidis* genome (E value: 0.0) (GenBank accession number AF269568) and contains 16S rRNA and 23S rRNA DNA sequences. Multiple regions of this DNA sequence match DNA sequences that map to human chromosome 5 (GenBank accession numbers AC009017 and AC008706) and human chromosome 16 (GenBank accession number AC009112), **b,** a prokaryote W DNA sequence (GenBank accession number AF286220) that matches a DNA sequence that maps to a Staphylococcus *aureus* nucleotide sequence, SRImec-I and SCCmec-I region (highest E value: 1e-168) (GenBank accession number AB033763) and contains the opening reading frame, ORFN042 (DDBJ accession number BAA94662.1), **c,** a prokaryote W DNA sequence (GenBank accession number AF286221) that matches a DNA sequence that maps to the Staphylococcus *epidermidis* genome (GenBank accession number AF269325) and **d,** a prokaryote W genomic sequence (GenBank accession number AF266551) shows similarity to the mRNA of a putative Hs1pro-1 homologue from *Glycine max* (soybean) (highest Expect value: 0.0) (GenBank accession number AF280812). This mRNA corresponds to less than half of an approximate three kilobase DNA sequence, the remainder of which is novel with no significant DNA sequence matches to DNA sequences contained in current nucleotide sequence databases (second figure).

© 2001 deNovo Biologic | www.denovo-bio.com

of human chromosomal DNA sequences[28,30]. The potentially large number of kilobase-sized human genome-related extragenic and genetic elements transferred 'vertically' in these experiments (as indicated by the *Alu*/LINE-1 PCR data) is at odds with assumptions regarding the rate of horizontal transfer and/or the maintenance of horizontally transferred sequences in bacterial genomes, both considered to be very low[29]. As corollary, these findings may indicate a eukaryotic origin for bacterial reverse transcriptase[31-33].

Some prokaryote W shotgun library DNA sequences match DNA sequences that have been mapped to human chromosomes 5, 9, 16 and 17. These data indicate that prokaryote-like DNA sequences, in particular Gram-positive eubacterium-like DNA sequences, may be 'hard-wired' in the human genome. This result is consistent with the suggestion that "…the presence of many bacterial genes with many kinds of functions…" within the eukaryotic nuclear genome "… should not be a surprise." [34]

Although a portion of the prokaryote W genomic shotgun library sampling contains DNA sequences that match DNA sequences that map to the *Staphylococcus epiderimidis* genome, most of the prokaryote W genomic shotgun library sampling (60%) consisted of DNA sequences with no significant matches by BLASTN analysis (matches with E values >1e-20), indicating a genetic/genomic novelty that may be due to widespread retroelement-mediated recombination events. Contrarily, as the human genome nucleotide sequence database expands, some novel DNA sequences may be found to match DNA sequences that map to the human genome. The plant-gene homologue discovered in the prokaryote W genome may be human genome-related since recent efforts to sequence the human genome reveal a novel gene similar to a plant gene within the human genome. Immunoblot studies indicate that these prokaryotes *express* human genes/proteins that map to many different human chromosomes and retroviral proteins. In addition, the variations in genome size and riboprinting between prokaryote W colony variants emphasize the highly plastic and pleomorphic nature of prokaryotes derived with this bioprocess as might be expected of a continuing process of eukaryote-to-prokaryote genome evolution.

Finally and crucially important, the rigorous, independently tested sterile cell culture methods used minimized the possibility that human genomic DNA sequences were transferred to microbes contaminating the cell culture systems. Altogether, this *in vitro* and *in silico* evidence is not easily reconcilable with a scenario of a xenologous eukaryote-to-prokaryote horizontal genomic DNA transfer but rather is more compatible with a scenario of orthologous eukaryote-to-prokaryote vertical genomic DNA transfer and a genome evolution associated with speciation. Clearly, the resulting genome is chimerical and rightly cannot be deemed to be 'staphylococcal' in any traditional taxonomic/genotypic sense.

Changes in the Oxygen Environment Drive Chromosomal Rearrangements and the Formation of Circular Double Minute Chromosomes in Tumor Cells

In the context of genome evolution, recent work by Coquelle and colleagues[35] might shed light on the role of an oxygen-related environmental pressure in eukaryote-to-prokaryote genomic DNA transfer. In tumor cells, they have demonstrated that hypoxia is a potent stimulus for fragile site induction, fusion of double minute chromosomes (DMs) and their targeted reintegration into chromosomal fragile sites, generating homogenously staining regions (HSRs). The HSRs so formed contain nonsyntenic sequences, *i.e.* sequences derived from different chromosomes, frequently observed in solid tumors. DMs are extrachromosomal circular DNA elements that contain SINEs and LINEs (including *Alu* and LINE-1 elements) in addition to other chromosomal elements that may be nonsyntenic. It has been postulated that the generation of *Alu*-containing extrachromosomal DNA such as DMs involves reverse transcription. Extrachromosomal DNA circularization may occur also by homologous or nonhomologous recombination[36]. In addition, anoxia has been found to be a potent stimulus for the induction of various retroelements[16,17]. Correspondingly, the genomes of these prokaryotes may be derived through hypoxia-driven formation and fusion of numerous, complex circular DMs, containing SINEs, LINEs and other genetic elements, which undergo selection to the level of a (circular) prokaryotic genome. This model of genome transformation is similar to the model of retroviral "plasmid transfer"-type linear-to-circular eukaryote-to-prokaryote genome transformation proposed by Poole and co-workers[8].

Human Retroelements Likely Play a Central Role in this Evolutionary Bioprocess and Might Account for the Genome Plasticity Observed in Prokaryote W Colony Variants

Retroviral and endogenous retrotransposons in the human eukaryotic cell genome(s), together with oxygen-related environmental pressure, likely act as driving forces in this *rapid* eukaryote-to-prokaryote genome evolution. Accumulation of retrosequences that give rise to novel genes may accelerate genomic change/evolution, catalyzing organismic evolution[37]. Genome reorganization and/or evolution may involve exon shuffling by L1 retrotransposition, which is postulated to represent a general mechanism for the evolution of new genes[38]. Other genetic/retrogenetic mechanisms of this genome evolution may include retrotransposon reverse-transcriptase-mediated repair of chromosomal breaks by non-homologous recombination[39,40].

Several prokaryotes isolated with this culture system exhibited pleomorphic colony variation in plate culture, similar to that observed in the prokaryote W colony variants. The genome sizes of two prokaryote W colony variants, observed during the pulse-field gel electrophoresis procedure for genome sizing, differed by approximately 500 kilobases. Riboprinting and shotgun library DNA sequencing of the prokaryote W colony variant with the smaller genome size of approximately 1.6 megabases indicates this colony variant belongs to the *Staphylococcus epidermidis* species group although documented genome sizes of this species group are no less than approximately 2.2 megabases[19].

Similar genome plasticity with significant variations in genome size up to the megabase range occurs in various species of streptomyces and appears to be related to genomic instability and recombination events associated with streptomyces chromosomal terminal inverted repeats[41,42]. This genomic instability has been related to streptomyces colony development[43].

In the case of the prokaryote W colony variation, the associated genome plasticity is likely due to recombination mediated by human repeats or retroelements, *i.e.* SINEs, LINEs, *etc.*, that are present in the prokaryote W genome(s). It should not go unnoticed that a short well-conserved region of the human eukaryotic *Alu* retroelement has homology with the prokaryotic *chi* nucleotide sequence that stimulates homologous recombination in various eubacterial species[44]. In addition, an *Alu* RNA motif is highly conserved in signal recognition particle RNAs from higher eukaryotes to yeast and from archaea to some Gram-positive eubacteria[45]. The bioprocess findings presented indicate that the human *Alu*-related nucleotide motifs found in prokaryotes in other investigations might have eukaryotic origins.

© 2001 deNovo Biologic | www.denovo-bio.com

Biocomputing Analogies Help Explain Eukaryote-to-Prokaryote Genome Evolution

One explanation for this genome evolution involves "the logic of soft-wired genomes" as described by Herbert and Rich[46,47]. Here, organisms generate RNA species that differ markedly from their DNA through the interaction of numerous gene products, variable RNA processing networks and the environment to determine RNA-species outcomes. RNA processing events implement mechanisms that are functionally equivalent to binary "logic gates." This combinatorial RNA-based 'logic' permits different sets of phenotypes or ribotypes to be selected from the DNA repertoire. These ribotypes are available for reverse transcription and integration into a DNA genome. The numerous human eukaryotic mRNA transcripts of a particular ribotype might co-opt the reverse transcription enzymatic machinery of human LINEs[48] and any transforming retrovirus, if present, to generate diverse functional processed pseudogenes/genes that serve the genetic functions of a prokaryotic organism. In this oxygen environment-driven genomic DNA transfer, a prokaryote-like ribotype 'soft-wired' genome may be generated from the DNA of multiple human chromosomes and selected for reverse transcription into a 'hard-wired' prokaryote-like DNA genome. Some human nucleotide sequences might pass from ribotype through the combinatorial RNA-based "logic gates" unscathed (as in Fig. 4a-e, Fig. 5a-c and Fig. 6d). Others might recombine to become novel DNA sequences or possibly eubacterial and plant-like sequences as observed in the prokaryote W shotgun library analysis.

From a population biology perspective, it is assumed that the selection of a 'successful' prokaryote-like ribotype, *ie.* one that is reverse transcribed into a functional prokaryotic genome, is a low-probability event. In some experiments, approximately one prokaryotic genome of approximately *two* megabases in size was derived from approximately 7.5×10^7 human eukaryotic nuclear genomes, comprising a collective genome-population size or 'genome space' of approximately 2.25×10^{11} megabases,. This represents a genome reduction in megabases by a factor of approximately 8.9×10^{-12} (human eukaryotic nuclear 'genome space' **:** prokaryotic genome). Clearly, in this bioprocess, there appears to be much human eukaryotic nuclear 'genome space' available for the biocomputing of a successful, albeit low-probability, prokaryote-like ribotype. Of course, this scenario does not take into account likely biocomputing contributions from human eukaryotic organellar 'genome space.'

Some aspects of the 'soft-wired' genome scenario of eukaryote-to-prokaryote genome evolution may have some resemblance to the mechanism of retroelement-related double minute chromosome formation noted above. Shapiro offers another biocomputing analogy for genome reorganization in which genomes, organized like integrated computer programs, are built upon a "system architecture" composed of repetitive DNA elements[49,50]. Altogether, this mechanism of genome evolution is likely inextricable from a higher order process of speciation.

This Bioprocess Might Provide Insights into Evolutionary Features of Tumorigenesis Which Include Eukaryote-to-Prokaryote Genome Evolution

It is possible that this work has relevance in studies of the role of highly pleomorphic staphylococci in tumorigenesis[51]. The pleomorphism of eubacteria with staphylococcal *features* isolated with this bioprocess likely is associated with a human retroelement-mediated genome plasticity. In this vein, it has been postulated that the evolutionary role of tumors is to provide the conditions for the expression of previously evolutionary new genes and, so, the material for the origin of new cell types[52]. These evolutionary new genes are dormant prior to their expression in a tumor. As regards the oxygen-related bioprocess described, the evolutionary role of retrovirally transformed eukaryotes, which possess tumor cell-like characteristics, is to provide the conditions for the expression of evolutionary new genes (prokaryote-like) and, so, the material for the origin of new cell types, *i.e.* prokaryotes. Accordingly, eukaryote-to-prokaryote genome evolution and speciation might be fundamental aspects of the evolutionary processes of tumorigenesis and immune deficiency in the multicellular eukaryotic organism. Similarly, recently documented constant genome reorganization in malignancy, attributed to chromosomal breakage-fusion-bridge events and indicated by genetic intratumor heterogeneity, appears to provide a flexible genetic system for clonal evolution and tumor progression[53]. The constant eukaryotic genome reorganization observed in malignancy might provide the combinatorial DNA 'shuffling' necessary to eukaryote-to-prokaryote genome evolution and speciation Correspondingly, it has been postulated that carcinogenesis involves *Alu*-related "re-evolution" from eukaryotic to prokaryotic replication units in eukaryotes[54,55].

Questions Remain

Several features of this eukaryote-to-prokaryote genome evolution and speciation remain to be determined. It is beyond the scope of the present work to speculate as to what, if any, prokaryotic genetic functions might be served by many of the numerous 'intact' human genomic DNA sequences found in the prokaryote W genome. However, as discovered in human chromosomes 5, 9, 16 and 17, other eubacterium-like DNA sequences may be 'hard-wired' in the human eukaryotic nuclear genome, perhaps endosymbiotic remnants[1-3]. These 'human/eubacterial' DNA sequences await a mobilizing environmental pressure to make the eukaryote-to-prokaryote genomic transition.

The largest portion of the prokaryote W genomic shotgun library sampling (60%) consisted of DNA sequences with no siginificant matches to DNA sequences as of December 2000 using NCBI GenBank nucleotide databases and BLAST programs. Subsequently, preliminary *in silico* analysis of these DNA sequences has revealed the presence of coding regions and translated nucleotide sequences with ill-defined, low homologies to proteins from a wide variety of taxa (human, prokaryotic and plant), indicating possible protein-related functions. Some of these DNA sequences might be found to match DNA sequences that map to the human genome as the human genome nucleotide sequence database expands and/or the methods of *in silico* analysis advance. Other 'unknown' DNA sequences, if generated through recombination and reverse transcription of human genome DNA elements, might be found to be chimerical, containing combinations of different DNA elements related to diverse genomes, human, prokaryote, plant and other.

Traditionally, prokaryotic genomes have been thought to lack repetitive DNA sequences, appearing to be optimized towards compactness. Recent analyses of complete prokaryotic genomes reveal unexpectedly high numbers of repetitive DNA sequences[56-59]. Whether or not these repetitive DNA sequences found in other prokaryotes have any direct relationship to those observed in the prokaryotes examined in the present study remains to be determined. However, it is likely that in both cases the repetitive DNA sequences participate in various mechanisms of prokaryotic genome plasticity, including amplifications, deletions, inversions, translocations and transpositions..

© 2001 deNovo Biologic | www.denovo-bio.com

Many questions beyond those concerning genome evolution alone remain. A genetic 'dissection' of bioprocess products, *i.e.* prokaryotes, was instrumental in determining prokaryotic origins. However, it is unlikely that current paradigms of evolutionary analysis, which largely focus on reductive genetic or molecular mechanisms that are often assumed to occur in the distant past or over expanses of time not directly observable, can give an adequate explanation for the coordinated transformation of numerous cellular systems, *i.e.* those biochemical, genetic, structural, *etc.*, necessary to a *rapid* eukaryote-to-prokaryote cellular evolution. In this emergent phenomenon of speciation, "the whole is greater than the sum of its parts" where the lowest *living* level of reduction, clearly not molecular, is at the level of the organism.

Using various models of evolutionary reconstruction, a small group of evolutionary biologists has proposed recently a eukaryotic rooting for the universal tree of life, postulating a eukaryotic origin for prokaryotic life-forms[7-13, 60]. The experiments and findings presented offer empirical evidence for a present-day, oxygen environment-driven, eukaryote-to-prokaryote vertical genomic DNA transfer and genome evolution of speciation. These studies suggest that some prokaryotic lineages are polyphyletic, the result of discrete, contemporaneous events of eukaryote-to-prokaryote cellular evolution. When this work is extended, it is likely that other novel evolutionary relationships will be discovered including some pertaining to the evolution of the mammalian retrovirus. Biotechnological applications need not await a more complete understanding of the mechanisms of this bioprocess, a type of directed reductive evolution, since its simplicity lends readily to the *rapid* generation of bacterial expression systems, human and animal biologicals, novel chimerical proteins and vaccines. □

Methods

Retrovirally Transformed Human Eukaryotic Cell Line
The retrovirally transformed human eukaryotic cell line was produced by cloning human cerebral microvascular endothelial cells that had been transformed by L cell-conditioned medium. Similar methods have been used to establish retrovirally transformed porcine cerebral microvascular endothelial cell lines[22]. This cell line is available from the American Type Culture Collection (ATCC) (Manassas, VA) and is designated as ATCC CRL 11655.

Sterility Testing of the Retrovirally Transformed Human Eukaryotic Cell Line, ATCC CRL 11655
Prior to use in the bioprocess experiments, ATCC CRL 11655 was found to be free of microbial contamination in in-house studies and in independent sterility testing for fungal, bacterial and mycoplasma contamination (BioReliance Corp., Rockville, MD) using stringent controls. Sterility testing also included a highly sensitive, double-step-PCR molecular phylogenetic method using nested, degenerate primers to detect gene sequences coding for the evolutionary conserved 16S rRNA of some twenty-five mycoplasmas, including those most commonly found in cell cultures[18].

Sterile Cell Culture System
Strict sterile procedures were followed during all antibiotic-free cell culture and cell culture system manipulations. For the *environmental pressure phase*, sterilely vented flasks containing approximately 1.8 x 10^8 ATCC CRL 11655 cells in nitrogen gas-flushed medium and a medium-only control flask were placed in a modified anaerobic jar. The jar was purged with sterile-filtered N_2 gas until the oxygen level read 0% on at least two consecutive analyses by oximetry (1-2 hr). The jar with gas lines clamped was placed in a standard cell culture incubator for 24 hours with temperatures of 30^0C ± 2^0C or 36^0C ± 2^0C. Control flasks containing an equal number of ATCC CRL 11655 cells were fed with the medium and incubated in air/5% CO_2 in the same incubator. At 24 and 48 hours, all flasks were taken from the incubator, uncapped and exposed to air for 2-3 minutes under a laminar flow hood. The flasks were returned to their respective conditions for the second and third 24-hour periods repeating in each cycle the steps outlined above. For the *propagation phase*, flasks were scraped sterilely to dislodge adherent eukaryotic cells. Flask contents, containing eukaryotic cells and media, were thoroughly mixed. Flask contents were either pooled or processed individually and transferred to aerobic bacteriological cell culture. In some experiments, flask contents were filtered through 0.22 µm sterile filters prior to transfer thus excluding eukaryotic cells from bacteriological cell culture in the propagation phase and possibly indicating a cell-wall deficient characteristic of the prokaryotes propagated. The propagation phase, lasting up to 15 to 21 days, included a wide range of sterility tests for bacteria (aerobic and anaerobic), fungi, and mycoplasmas with extensive use of positive and negative controls. Any microorganisms isolated in the propagation phase were subcultured.

Independent Testing of Sterile Cell Culture System
Experiments using this cell culture system were performed independently at BioReliance Corp. (Rockville, MD; formerly known as Microbiological Associates, Inc.) with sterility testing protocols that followed stringent U.S. FDA guidelines. A total of four bioprocess experiments or 'runs' were performed; two with alternating anaerobic/aerobic conditions during the environmental pressure phase and two with anaerobic conditions only during the environmental pressure phase. For these experiments, a total of *978* separate cultures using a wide variety of bacteriological cell culture media were made to test for bacterial, fungal and mycoplasma growth. These cultures included those for experimental cell samples, control cell samples, media control samples and assay positive and negative controls. All negative controls were negative. All positive controls were positive. For the two bioprocess experiments using alternating anaerobic/aerobic conditions during the environmental pressure phase, the experimental cell sample cultures totaled *151*. A full reporting of this validation study can be viewed at http://www.denovo-bio.com/pdf/6AE40510.pdf.

Prokaryotes Isolated Deposited at the ATCC
Prokaryotes P and W, described in this work, were deposited at the ATCC. These microorganisms can be obtained from the ATCC under the designations ATCC 55589 and ATCC 55590, respectively.

Genome Sizing
One-dimensional pulsed-field gel electrophoresis was used to size the prokaryote P and prokaryote W genomes[20].

Riboprinting
Riboprinting was performed using the Qualicon RiboPrinter® Microbial Characterization System (Qualicon, Inc., Wilmington DE).

Immunoblots
Standard immunoblotting methods were employed. For the retroviral *gag* protein, a 1:500 dilution of antisera raised against the Moloney MuLV *gag* p30 core protein and $Pr65^{gag}$ polyprotein was used [22]. Similar immunoblotting methods were employed to detect human proteins.

© 2001 deNovo Biologic | www.denovo-bio.com

Genomic DNA Preparation

Eukaryotic genomic DNA was isolated using standard methods. Prokaryotic genomic DNA was isolated with a protocol used by the ATCC for Gram-positive prokaryote genomic DNA preparations. Prokaryotic genomic DNA was prepared only from prokaryotes propagated in bacteriological liquid media used for scale-up culture. These prokaryotes ('P' and 'W') were isolated initially from individual bacterial colonies on plate culture during the propagation phase after flask contents from the environmental pressure phase had been filtered through sterile 0.22 μm filters prior to transfer. These culture methods excluded the unlikely possibility that eukaryotic cells from the environmental pressure phase might propagate in and contaminate bacteriological cell cultures in the propagation phase.

Human Alu/LINE-1-directed Polymerase Chain Reaction

Oligonucleotide primers 5' CGACCTCGAGATC-TYRGCTCACTGCAA 3' ('517') and 5 ' AAGTCGCGGCCGC-TTGCAGTGAGCCGAGAT 3' ('TC-65'), human *Alu*-specific[23], and 5' ACGTTGTGCACATGTACC 3' ('L1S), human LINE-1-specific[24], were synthesized. These 'human-specific' primers were paired and used in PCR conditions which included the use of 1X *Taq* polymerase buffer (BRL), 1.5 mM $MgCl_2$, 200 mM dNTPs (Perkin Elmer), 2.5 units *Taq* polymerase and 2 ug genomic DNA. Cycling parameters were as follows: 35 cycles of 94^0C denaturation (1 min), 55^0C annealing (45 sec), 70^0C extension (5 min), and a final extension at 72^0C (10 min) in an automated thermal cycler (Perkin-Elmer). PCR products were electrophoresed on a 0.8% agarose gel.

DNA Sequencing and In Silico Analysis

Human Alu/LINE-1-directed PCR products amplified from prokaryotic genomic DNA of isolates P and W were cloned using a T/A cloning method. Sequencing was performed by primer extension. A genomic shotgun library of prokaryote W was prepared and sequenced by primer extension. Genomic fragments were cloned using a double-adaptor method (SeqWright, Houston, TX). Sequencing reactions were performed on clones randomly selected from this library. Bioinformatics analysis of DNA sequences relied primarily upon the use of the NCBI GenBank nucleotide databases (current to December 2000) and BLAST programs[25] for sequence match searches [www.ncbi.nlm.nih.gov/BLAST]. In general, sequence matches with E values > 1e-20 were considered insignificant for the purposes of this study.

References

1. Margulis, L. *Symbiosis in cell evolution* (W. H. Freeman, New York 1981, 2nd edition 1993).
2. Sapp, J. *Evolution by association: a history of symbiosis* (Oxford Univ Press, New York, 1994).
3. Martin, W. & Muller, M. The hydrogen hypothesis for the first eukaryote. *Nature* **392,** 37-41 (1998).
4. Gupta, R.S. Protein phylogenies and signature sequences: evolutionary relationships within prokaryotes and between prokaryotes and eukaryotes. *Antonie Van Leeuwenhoek* **72,** 49-61 (1997).
5. Gupta, R.S. What are archaebacteria: life's third domain or monoderm prokaryotes related to gram-positive bacteria? A new proposal for the classification of prokaryotic organisms. *Mol. Microbiol*. **29,** 695-707 (1998).
6. Doolittle, W.F. Phylogenetic classification and the universal tree. *Science* **284,** 2124-2149 (1999).
7. Forterre, P. Thermoreduction, a hypothesis for the origin of prokaryotes. *C. R.Acad. Sci .Paris Life Sci* . **318,** 415- 422 (1995).
8. Poole, A.M., Jeffares, D.C. & Penny, D. The path from the RNA world. *J. Mol. Evol*. **46,** 1-17 (1998).
9. Jeffares, D.C., Poole, A.M. & Penny, D. Relics from the RNA world. *J. Mol. Evol.* **46,** 18-36 (1998).
10. Lopez, P., Forterre, P. & Phillipe, H. The root of the tree of life in light of the covarion model. *J. Mol. Evol.* **49,** 496-508 (1999).
11. Phillipe, H. & Forterre, P. The rooting of the universal tree of life is not reliable. *J .Mol. Evol.* **49,** 509-523 (1999).
12. Forterre P, & Phillipe H: Where is the root of the universal tree of life? *Bioessays* 1999, 21: 871-879.
13. Poole AM, Jeffares DC & Penny D: Early Evolution: prokaryotes, the new kids on the block. *Bioessays* 1999, 21: 810-889.
14. Eldredge N & Gould SJ: Punctuated equilibria: an alternative to phyletic gradualism. In *Models in Paleobiology*. Edited by Schopf TJM. San Francisco: W.H. Freeman, 1972, 82-115.
15. McDonald, J.F. Transposable elements: possible catalysts of organismic evolution. *Trends Ecol. Evol.* **10,** 123-126 (1995).
16. Anderson, G.R. & Stoler, D.L. Anoxia, wound healing, VL30 elements, and the molecular basis of malignant conversion. *BioEssays* **15,** 265-271 (1993).
17. Estes, S.D., Stoler, D.L. & Anderson, G.R. Anoxic induction of a sarcoma virus-related VL30 retrotransposon is mediated by a cis-acting element which binds hypoxia-inducible factor 1 and an anoxia-inducible factor. *J. Virol.* **69,** 6335-6341 (1995).
18. Hopert, A., Uphoff, C.C., Wirth, M., Hauser, H. & Drexler H,G. Specificity and sensitivity of polymerase chain reaction (PCR) in comparison with other methods for the detection of mycoplasma contamination in cell lines. *J. Immunol. Meth.* **164,** 91-100 (1993).
19. George, C.G. & Kloos, W.E. Comparison of the SmaI-digested chromosomes of Staphylococcus epidermidis and the closely related species *Staphylococcus capitis* and *Staphylococcus caprae*. *Int. J. Syst. Bacteriol.* **44,** 404-409 (1994).
20. Romling U, Schmidt K & Tummler B: One-dimensional pulsed-field gel electrophoresis. In *Bacterial Genomes.* Edited by de Bruijn FJ, Lupski JR, Weinstock, GM. New York. Chapman and Hall, 1998, 312-325.
21. Shinnick, T.M., Lerner, R.A. & Sutcliffe J.G. Nucleotide sequence of the Moloney murine leukemia virus. *Nature* **293,** 543-548 (1981).
22. Robinson, D.H., Warren, M.K., Liang, L-T., Oprandy, J.J., Nielsen, T. B. & Kang Y-H. Retroviral transformation of cerebral microvascular endothelial cells: macrophage-like and microvascular endothelial cell properties. *Blood* **77,** 294-305 (1991).
23. Nelson,D.L., Ledbetter, S.A,. Corbo, L., Victoria, M.F., Ramirez-Solis, R., Webster, T.D., Ledbetter, D.H. & Caskey, C.T. Alu polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. *Proc. Natl. Acad. Sci. USA* **86,** 6686-6690 (1989)
24. Brooks-Wilson, A.R., Smailus, D.E., Weier, H.U. & Goodfellow, P.J. Human repeat element-mediated PCR: cloning and mapping of chromosome 10 DNA markers. *Genomics* **13,** 409-414 (1992).
25. Altschul, S.F., Madden, T.L., Schaffer A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389-3402 (1997).
26. Gupta, R.S. The natural evolutionary relationship among prokaryotes. *Crit. Rev. Microbiol.* **26,** 111-131 (2000).
27. Production Sequencing Facility of the DOE Joint Genome Institute [http://www.jgi.doe.gov/]
28. Hattori, M. *et al.* The DNA sequence of human chromo-some 21. *Nature* **405,** 311-319 (2000).
29. Ochman, H., Lawrence, J.G. & Groisman E.A. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405,** 299-304 (2000).
30. Dunham, I. *et al.* The DNA sequence of human chromosome 22. *Nature* **402,** 489-495 (1999).
31. Inouye, S. & Inouye, M. Structure, function, and the evolution of bacterial reverse transcriptase. *Virus Genes* **11,** 81-94 (1995).
32. Rice, S.A. & Lampson B.C. Bacterial reverse transcriptase and msDNA *Virus Genes* **11,** 95-104 (1995).
33. Lampson, B.C. & Rice, S.A. Repetitive sequences found in the chromosome of the myxobacterium *Nannocystis exedens* are similar to msDNA: a possible retrotransposition event in bacteria. *Mol. Microbiol.* **23,** 813-823 (1997).
34. Doolittle, W.F. You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends Genet*. **14,** 307-311 (1998).

© 2001 deNovo Biologic | www.denovo-bio.com

35. Coquelle, A,. Toledo, F., Stern ,S., Bieth, A. & Debatisse M. A new role for hypoxia in tumor progression: induction of fragile site triggering genomic rearrangements and formation of complex DMs and HSRs. *Mol. Cell* **2,** 259-265 (1998).
36. Abken, H. Extrachromosomal DNA elements in humor tumor cells. *The Cancer Journal* **8** [http://www.infobiogen.fr/agora/journals/cancer/articles/8-3/abken.html] (1995).
37. Brosius, J. RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. *Gene* **238,** 115-134 (1998).
38. Moran, J.V., DeBerardinis, R.J. & Kazazian, H.H. Exon shuffling by L1 retroransposition. *Science* **283,** 1530-1534 (1999).
39. Teng, S-C., Kin, B & Gabriel, A. Retrotransposon reverse transcriptase mediated repair of chromosomal breaks. *Nature* **383,** 641-644 (1996).
40. Moore,J.K. & Haber J.E. Capture of retrotransposon DNA at the sites of double-strand breaks. *Nature* **383,** 644-646 (1996).
41. Fischer, G., Kyriacou, A., Decaris, B. & Leblond, P: Genetic instability and its possible evolutionary implications on the chromosomal structure of *Streptomyces*. *Biochimie* **79,** 555-558 (1997).
42. Fischer, G., Wenner, T., Decaris., B. & Leblond P. Chromosomal arm replacement generates a high level of intraspecific polymorphism in the terminal inverted repeats of the linear chromosomal DNA of *Streptomyces ambofaciens*. *Proc. Natl. Acad. Sci.* USA **95,** 14296-14301 (1998).
43. Dary, A., Martin, P., Wenner, T., Decaris, B. & Leblond P: DNA rearrangements at the extremities of the *Streptomyces ambofaciens* linear chromosome: evidence for developmental control. *Biochimie* **82,** 29-34 (2000).
44. Rudiger, N.S., Gregersen, N. & Kielland-Brandt, M.C. One short well conserved region of Alu-sequences is involved in human gene rearrangements and has homology with prokaryotic *chi*. *Nucleic Acids Res* **23,** 256-260 (1995).
45. Strub, K,. Moss, J. & Walter, P. Binding sites of the 9- and 14-kilodalton heterodimeric protein subunit of the signal recognition particle (SRP) are contained exclusively in the Alu domain of SRP RNA and contain a sequence motif that is conserved in evolution. *Mol. Cell. Biol* . **11,** 3949-3959 (1991).
46. Herbert, A. & Rich , A. RNA processing and the evolution of eukaryotes. *Nat. Genet.* **21,** 265-269 (1998).
47. Herbert, A. & Rich, A. RNA processing in evolution: the logic of the soft-wired genome. *Ann. N.Y. Acad.. Sci.* **870,** 119-132 (1999).
48. Esnault, C., Maestre, J. & Heidmann, T. Human LINE retrotransposons generate processed pseudogenes. *Nat. Genet.* **24,** 363-367 (2000).
49. Shapiro, J.A. Transposable elements as the key to a 21st century view of evolution. *Genetica* **107,** 171-179 (1999).
50. Shapiro, J.A. Genome system architecture and natural genetic engineering in evolution. *Ann. N .Y. Acad. Sci.* **870,** 23-35 (1999).
51. Wainwright, M. Highly pleomorphic staphylococci as a cause of cancer. *Med. Hypotheses* **54,** 91-94 (2000).
52. Kozlov, A.P. Gene competition and the possible evolutionary role of tumours. *Med. Hypotheses* **46,** 81-84 (1996).
53. Gisselsson, D. *et al.* Chromosomal breakage-fusion-bridge events cause genetic intratumor heterogeneity. *Proc. Natl. Acad. Sci.USA* **97,** 5357-5362 (2000).
54. Grossgebauer, K. Cancer: re-evolution of eucaryotic to procaryotic replication units – a result of genetic transposition? *Biosystems* **16,** 253-258. (1984).
55. Grossgebauer, K. Cancer: involvement of replicative origins? *Med. Hypotheses* **21,** 127-140 (1986).
56. Rocha, E.P., Danchin, A. & Viari, A. Analysis of long repeats in bacterial genomes reveals alternative evolutionary mechanisms in *Bacillus subtilis* and other competent prokaryotes. *Mol. Biol. Evol.* **16,** 1219-1230 (1999).
57. Rocha, E.P., Danchin, A. & Viari, A. Functional and evolutionary roles of long repeats in prokaryotes. *Res. Micriobiol.* **150,** 725-733 (1999).
58. Romero, D., Martinez-Salazar, J., Ortiz, E., Rodriguez, C. & Valencia-Morales, E. Repeated sequences in bacterial chromosomes and plasmids: a glimpse from sequenced genomes. *Res. Micriobiol.* **150,** 735-743 (1999).
59. Metzgar, D., Thomas, E., Davis, C., Field, D. & Willis, C. The microsatellites of *Escherichia coli*: rapidly evolving repetitive DNA in a non-pathogenic prokaryote. *Mol. Microbiol.* **39,** 183-190 (2001).
60. Glansdorff, N. MicroReview: about the last common ancestor, the universal life-tree and lateral gene transfer: a reappraisal. *Mol. Microbiol.* **38,** 177-185 (2000).

Acknowledgments

I thank the following individuals for their technical, advisory and administrative contributions to this work: D.W. Wagner, M.P. Malven, E. Yu, E. Guingab, the late J.-H. Kang, R. Williams, L. Lawson, P. C. Robinson and several anonymous reviewers. The assistance of other individuals at the Naval Medical Research Institute (Bethesda, Maryland), the American Type Culture Collection (Manassas, VA), BioReliance Corp. (Rockville, MD), ARUP Laboratories (Salt Lake City, UT) and SeqWright, Inc. (Houston, TX) is appreciated greatly. This work was supported in part by a grant from the Office of Naval Research. The opinions and assertions contained herein are the private ones of the author and are not to be construed as official or reflecting the views of the Navy Department or the Naval Service at large.

Postscript

Two versions of the human genome draft sequence and its analysis were published[1,2] shortly after the completion of this research article. The following observations follow from perspectives gained from initial reviews of these publications and their possible relationships to the present work.

Interestingly, close to half of the human genome appears to consist of various transposable elements including LINEs, SINEs, LTR retrotransposons and DNA transposons. Transposons appear to be a creative evolutionary force, responsible for important innovations in genome evolution, human and other. At least 47 human genes may be human transposon-derived, many newly recognized through human genome sequencing and analysis. The majority ofthese human genes are derived from DNA transposons. Evidence strongly supports the role of LINE-1 activity in the origin of at least eight human genes[3].

Surprisingly, hundreds to thousands of sequences in the draft human genome sequence appear to have been reverse transcribed from various noncoding RNAs and coding RNAs related to the translation system.[1] As much as 40% of retrotransposed intronless paralogues are related to ribosomal proteins with another 10% of retrotransposed intronless paralogues related to translation elongation factors[2]. The large number of human pseudogenes, fragments and paralogues, retrotransposed from components of the translation system, suggests an active evolutionary relationship between reverse transcription and the translation system. This finding suggests a primary role for reverse transcription in the translation system transformation associated with eukaryote-to-prokaryote cellular evolution. (See Fig. 6a.)

Human proteome analysis of the human genome draft sequence indicates the presence of 223 proteins that have significant similarity to proteins from bacteria with no comparable similarity to other non-vertebrate eukaryotic proteins[1]. These 'bacterial' proteins comprise perhaps less than one percent of predicted human proteins. However, approximately one fifth of human proteins are homologous to both eukaryotic and prokaryotic proteins.

© 2001 deNovo Biologic | www.denovo-bio.com

The authors of the international consortium report[1] favor a mechanism of prokaryote-to-eukaryote horizontal gene transfer to account for these 'bacterial' proteins in the human proteome repertoire. However, they "...cannot formally exclude the possibility that gene transfer occurred in the opposite direction - that is, that the genes were invented in the vertebrate lineage and then transferred to bacteria." They regard this scenario as less likely due to restraints on prokaryote-to-prokaryote horizontal dissemination after initial eukaryote-to-prokaryote horizontal gene transfer (likely presumed to involve one or a few prokaryotes). In addition, the authors consider the functional repertoire of these proteins (primarily intracellular enzyme homologues) to be "...uncharacteristic of vertebrate-specific evolutionary innovation..." which appear to be extracellular protein-centric. Not surprisingly, the authors do not include consideration of multiple events of vertical DNA transfer associated with eukaryote-to-prokaryote genome evolution and speciation.

Another observation concerns the methods of analysis of the human genome sequence in order to mine these 'bacterial' proteins. 'Bioinformatically,' the human genome sequence was filtered to "…eliminate sequences that were essentially identical to known bacterial plasmid, transposon or chromosomal DNA (such as the host strains for large-insert clones)."[1] Although neither report contains specifics regarding exactly which bacterial species were included for sequence editing, this method might have had an impact on the genome analysis of the present work. Pertinently, the Celera team masked approximately 35 megabases of 'nonhuman' DNA sequence from the GenBank data it used in constructing its version of the human genome draft sequence[2].

In early htgs database searches and sequence alignments in the present work, two reported prokaryote W DNA sequences (GenBank accession numbers AF286220 and AF286221; Fig. 6b-c) matched both staphylococcal genome sequences and human chromosome 5 DNA sequences (GenBank accession number AC022126, gi version 8576127). The translated BLAST search (translated nucleotide queried against a protein database) indicated that one of these 'staphylococcal/human' DNA sequences (AF286221) was homologous to a portion of a bacterial copper-transporting ATPase. As of November 2000, these prokaryote W DNA sequences no longer matched revisions of this human chromosome 5 DNA sequence (GenBank gi versions subsequent to gi 8576127).

Contamination of human genome sequences by DNA sequences from a bacterial host strain used in human genome sequencing such as *E. coli* is not uncommon. However, it is less likely that staphylococcal genome sequences contaminated the human genome sequences during human genome cloning and sequencing efforts. Therefore, it is reasonable to speculate that these prokaryote W DNA sequences indeed have matches to both staphylococcal genome sequences and human genome sequences and that the matches to human genome sequences were edited *in silico* in human genome draft sequence revisions. If so, human genomes might contain many bacterial DNA sequences, staphylococcal and other, that have been and will be omitted due to similar bioinformatics sequence-filtering biases. If such bacterial DNA sequences were incorporated into a *specific* human genome *in vivo*, *i.e.* prior to its cloning and sequencing, it is reasonable to argue that *as a part of that human genome*, these 'human/bacterial' DNA sequences should remain in the human genome sequence record(s). It is likely that individual human genomes vary widely with respect to the number and types of 'human/ bacterial' DNA sequences they contain.

As corollary, preliminary *in silico* analysis of the prokaryote W genomic shotgun library in April 2001, using the EMBL nucleotide databases and the European Bioinformatics Institute WU-BLAST 2.0 program, indicates the presence of several DNA sequences that appear to be chimeras of staphylococcus genome-related DNA sequences and human genome-related DNA sequences (see Appendix II for representative chimerical DNA sequence alignments). Many other shotgun library DNA sequences appear to be human genome-related, fungus genome-related (*Botrytis cinerea*) and plant genome-related. The phylogenetic diversity of these DNA sequences supports a mechanism of eukaryote-to-prokaryote genome evolution that involves widespread genetic recombination and possibly has some relationship to the pleomorphism observed in prokaryotes derived with this oxygen-related bioprocess. As recently as December 2000, these prokaryote W genomic shotgun library DNA sequences were found to have no significant matches to DNA sequences contained in the NCBI GenBank nucleotide databases using the NCBI BLASTN program.

It is possible that other 'nonhuman' sequences, such the plant-like sequence in Fig. 6d, have been and will be filtered out as sequencing and *in silico* analysis of the human genome continues. In this vein, ongoing *in silico* analysis of the prokaryote W genome shotgun library using a commercial cDNA database indicates the presence of cDNA sequence fragments that show significant similarities to human/plant eukaryotic elongation translation initiation factors, plant eukaryotic 60S ribosomal proteins and a plant eukaryotic arginine-tRNA-protein transferase. While 48-60% of genes involved in *Arabidopsis* protein synthesis have been found to have counterparts in other eukaryotic genomes, at least 36 *Arabidopsis* genes have significant similarities to human disease genes[4]. Accordingly, plant-like cDNA sequences, similar to that depicted in Fig. 6d, might be expected to be present in the human genome and, so, a human genome-derived prokaryotic genome.

Nonetheless, upon initial review of the human genome draft sequence and its analysis, it is clear that the human genome contains many potential genome-sculpting transposable elements along with many bacterial gene homologues, some recognized and some possibly relegated with other 'nonhuman' homologues to the 'cutting room floor' *in silico.* One can only guess at what other surprises will be discovered as human genome sequencing and its analysis continues. With the above observations in mind, it seems evermore credible that the human eukaryotic genome and cell might have the potential for grand evolutionary innovation, *i.e.* eukaryote-to-prokaryote genome evolution and speciation, if given the right environmental nudge by Nature.

dhr

1. Lander, E.S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409,** 860-921 (2001).
2. Venter, J.C. *et al.* The sequence of the human genome. *Science* **291,** 1304-1351 (2001).
3. Brosius, J. Genomes were forged by massive bombardments with retroelements and retrosequences. *Genetica* **107,** 209-238 (1999).
4. The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796-815 (2000).

Correspondence and requests for materials may be addressed to
e-mail: douglas.robinson@denovo-bio.com

Appendix I

A Representative Alu/LINE-1-directed PCR DNA Sequence Alignment, Fig. 4e

Sequence alignment for prokaryote W DNA sequence (GenBank accession number AF266544) that matches a DNA sequence that maps to human chromosome X (E value: 0.0) (EMBL accession number HSU227D1).

```
>gi|1171570|emb|Z68908.1|HSU227D1 Human DNA sequence from cosmid U227D1, between markers DXS366 and DXS87 on chromo-
some X     Length = 33667

 Score = 3808 bits (1921), Expect = 0.0
 Identities = 1955/1971 (99%)
 Strand = Plus / Plus

Query: 1     gaacctaaaattaaaattaaaaataatatatttgctagaaaattatgactgctctgttga 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 16154 gaacctaaaattaaaattaaaaataatatatttgctagaaaattatgactgctctgttga 16213

Query: 61    gtgatttgtagtcataaacaccattttgatctttaattccaaaaaacacaagtcacctag 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Sbjct: 16214 gtgatttgtagtcataaacaccattttgatctttaattccaaaaaacacaagtcacctat 16273

Query: 121   gtttttcaggagctgataattaagacacctcatatatattaatagatgaaccaaggatca 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 16274 gtttttcaggagctgataattaagacacctcatatatattaatagatgaaccaaggatca 16333

Query: 181   tttttcaaatcctctctccagattctttctgcagcctgtgaagagctgtttttggaaact 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 16334 tttttcaaatcctctctccagattctttctgcagcctgtgaagagctgtttttggaaact 16393

Query: 241   gcattagaagaaggaaaggcttggtttggaagaacaggatcagctttccaccctgtgcct 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 16394 gcattagaagaaggaaaggcttggtttggaagaacaggatcagctttccaccctgtgcct 16453

Query: 301   ggcagaggctatggctgccctccatgtgaagcttcaaattaaaatgaagttgtggtgcct 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 16454 ggcagaggctatggctgccctccatgtgaagcttcaaattaaaatgaagttgtggtgcct 16513

Query: 361   acgcagcttctggaactacaggggggttcttccacagcaggttacacagagctaaaataat 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 16514 acgcagcttctggaactacaggggggttcttccacagcaggttacacagagctaaaataat 16573

Query: 421   tgagtgtggtgttgtttaccaacatcagggagaaaaacagcaaaagtcaatataaaccat 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 16574 tgagtgtggtgttgtttaccaacatcagggagaaaaacagcaaaagtcaatataaaccat 16633

Query: 481   gaaagtagtaaattcccaattttgtcttggttaggtacatgttgggagagaagagaaaaa 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 16634 gaaagtagtaaattcccaattttgtcttggttaggtacatgttgggagagaagagaaaaa 16693

Query: 541   gagaaactacaactttagtcaggaaaaataaaaacaatttatagaaaaactaaccataag 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 16694 gagaaactacaactttagtcaggaaaaataaaaacaatttatagaaaaactaaccataag 16753

Query: 601   atgtggtctcccaaatgcagttaagaaaattcttatacttgtctggttctgtgatggtac 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 16754 atgtggtctcccaaatgcagttaagaaaattcttatacttgtctggttctgtgatggtac 16813

Query: 661   ataaatgtaaagggcttgagattttagtatgagagactcagattttagaaactccattcc 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 16814 ataaatgtaaagggcttgagattttagtatgagagactcagattttagaaactccattcc 16873

Query: 721   accaacccaactttaagcccttcctgctctcaaatgggaagcagaagaggttaagccaag 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 16874 accaacccaactttaagcccttcctgctctcaaatgggaagcagaagaggttaagccaag 16933

Query: 781   ccaatcttaaatgtcagtaaatgtaaatacaaatatttttggagtagctaaaatacaatc 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 16934 ccaatcttaaatgtcagtaaatgtaaatacaaatatttttggagtagctaaaatacaatc 16993

Query: 841   taccttcagataatgaaatgtttgtgtttaataacatacctttacatatgcatagtgctg 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 16994 taccttcagataatgaaatgtttgtgtttaataacatacctttacatatgcatagtgctg 17053

Query: 901   tttacttttcaaatcactttaacccattttacccttacaacaaagttaaagtgcgtatgg 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 17054 tttacttttcaaatcactttaacccattttacccttacaacaaagttaaagtgcgtatgg 17113
```


© 2001 deNovo Biologic | www.denovo-bio.com

```
Query: 961   tagttattttacagtttcaggatgagttaaattaaggcaccattgagttaatgatttgta 1020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 17114 tagttattttacagtttcaggatgagttaaattaaggcaccattgagttaatgatttgta 17173

Query: 1081  aagcagtcatgtttccaattttatgctgagtgccagtaagtcatagggatattccctgca 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 17234 aagcagtcatgtttccaattttatgctgagtgccagtaagtcatagggatattccctgca 17293

Query: 1141  tttataaccaggtgtccaaaaagaagatgagaattacttctgcttctccaatttcccatc 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 17294 tttataaccaggtgtccaaaaagaagatgagaattacttctgcttctccaatttcccatc 17353

Query: 1201  taaaatatttcatgattgaaaaatatatgtatttaagaacctactccagagctaaactgc 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 17354 taaaatatttcatgattgaaaaatatatgtatttaagaacctactccagagctaaactgc 17413

Query: 1261  agggttgtcaccatgtagacccatgatttgctccttctttggccaaaggcaagagaaaga 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 17414 agggttgtcaccatgtagacccatgatttgctccttctttggccaaaggcaagagaaaga 17473

Query: 1321  aaagagaagttgacagggtacccattttcatctcacccagaaggtagctggtcgttgttg 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 17474 aaagagaagttgacagggtacccattttcatctcacccagaaggtagctggtcgttgttg 17533

Query: 1381  acagtataactctccaaagaaacaggaggcgcactcttcaggaaataacccaagcaattc 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 17534 acagtataactctccaaagaaacaggaggcgcactcttcaggaaataacccaagcaattc 17593

Query: 1441  agcccctctgcaaagaacatctgttcctggccatacaagatcctttgcagctttnnnnnn 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 17594 agcccctctgcaaagaacatctgttcctggccatacaagatcctttgcagctttaaaaaa 17653

Query: 1501  ngcaaggctaggagaaaatgttgcagagggcttctcctgtgagggcagaaagcagaggga 1560
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 17654 agcaaggctaggagaaaatgttgcagagggcttctcctgtgagggcagaaagcagaggga 17713

Query: 1561  cagatggggagcagtgtctggagaaatcagggtcatgggggtgttagcatttttggggga 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 17714 cagatggggagcagtgtctggagaaatcagggtcatgggggtgttagcatttttggggga 17773

Query: 1621  ttagggctcagagcacattactgcattggacaaaagccaggagaaaagcagagagcaggc 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 17774 ttagggctcagagcacattactgcattggacaaaagccaggagaaaagcagagagcaggc 17833

Query: 1681  ttctcactaatgtatggagatagaaggccactaaaaatcttcaatttgcttctcataagg 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 17834 ttctcactaatgtatggagatagaaggccactaaaaatcttcaatttgcttctcataagg 17893

Query: 1741  caaccatattcaaggccccccttctttacacctttacctcactactgtcaaaactctaacc 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 17894 caaccatattcaaggccccccttctttacacctttacctcactactgtcaaaactctaacc 17953

Query: 1801  tctaccacaaaacctatactcctacagatttttcccttaatgccatgatgcaaaaatcc 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 17954 tctaccacaaaacctatactcctacagatttttcccttaatgccatgatgcaaaaatcc 18013

Query: 1861  aattagtttccatgatcttagtttagagcttggaacccagtgtttcctccttttcaccat 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 18014 aattagtttccatgatcttagtttagagcttggaacccagtgtttcctccttttcaccat 18073

Query: 1921  acaagccatgctgactttatttatcaaagcctaatggcatgttaatatttc 1971
             ||||||||||||||||||||||||||||||||| |||||||||||||||||
Sbjct: 18074 acaagccatgctgactttatttatcaaagcctagtggcatgttaatatttc 18124
```

Appendix II

Representative Chimerical DNA Sequence Alignments

Sequence alignments of a prokaryote W genomic shotgun library DNA sequence that appears to be a chimera of a staphylococcus genome-related DNA sequence and a human genome-related DNA sequence. The alignments were generated using the EMBL nucleotide databases and the European Bioinformatics Institute WU-BLAST 2.0 program.

```
EM_PRO:AF269879 AF269879 Staphylococcus epidermidis strain SR1 clone
            step.1028b05 genomic sequence.
        Length = 3609

  Plus Strand HSPs:

 Score = 1028 (160.3 bits), Expect = 2.6e-56, Sum P(2) = 2.6e-56
 Identities = 220/238 (92%), Positives = 220/238 (92%), Strand = Plus /  Plus

Query:   272 AGCCAGTAATTATGTCCCTGCTTTTTCTTTATTTATATAACTTTCATATTGAACGACTCG 331
             ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1886 AGCTAGTAATTATGTCCCTGCTTTTTCTTTATTTATATAACTTTCATATTGAACGACTCG 1945

Query:   332 TCATCAAAAGTTACCGTTTCAGCCTGTATTTGGATTCGCGCACCTTTCACAGAAATATC 391
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1946 TCATCAAAAGTTACCGTTTCAGCCTGTATTTGGATTCGCGCACCTTTCACAGAAATATC 2005


Query:   392 GTAATAACACTACGTATTAACATATCATCTTATTTTTACATGTTTTACAAAAGATATTTT 451
             |||| |||||||||||||||||||| |||||||||||| |||||||| ||||||||||||
Sbjct:  2006 GTAAAAACACTACGTATTAACATATAATCTTATTTTTAAATGTTTTAAAAAAGATATTTT 2065

Query:   452 TTATTTAGTTCCAGATTAGTCTTCGTCTCCTTTCGCTTATTGTTTCTTTTAGACGTTT 509
             |||||||||   |||||||| |  ||   |||| | |||||||| |||||||| ||||
Sbjct:  2066 TTATTTAGTAAAAGATTAGTATAAGTTAACTTTTGATTATTGTTACTTTTAGAAGTTT 2123



EM_HUM:AC018634 AC018634 Human Chromosome 7 clone RP11-243E12, complete
            sequence.
        Length = 174,241

  Plus Strand HSPs:

 Score = 1170 (181.6 bits), Expect = 7.2e-50, Sum P(2) = 7.2e-50
 Identities = 554/829 (66%), Positives = 554/829 (66%), Strand = Plus / Plus

Query:    416 TCATCTTATTTTTACATGTTTTACAAAAGATATTTT-TTATT-TAGTTCCAGATTAGTCT 473
              || |||| || || | | || | |      |  ||  || || |  |||    ||  |||
Sbjct: 144951 TCTTCTTCTTCTT-CTTCTTCTTCTTCTTCTTCTTCCTTCTTCTTCTTCTTC-TTCTTCT 145008

Query:    474 TCGTCTCCTTTCGCTTATTGTT-TCTT-TTAGACGT-TTCTCCTTAGCTTT-TTCTCCCC 529
              || ||| |||   ||| || || |||| ||   | | |||| |||  |||| ||||| |
Sbjct: 145009 TCTTCTTCTTCTTCTTCTTCTTCTCTTCTTCTTCTTCTTCTTCTT--CTTTCTTCTCTCT 145066

Query:    530 -CCTCTTCTTCC-CTT-TTC-TCCTCC--CTTCTTTCCTCATTCTTTTCCGCTTTCTTAT 583
               ||||||| ||  ||| ||| || |||  ||||||| ||  ||||||||  ||| ||| |
Sbjct: 145067 TCCTCTTCCTCTTCTTCTTCGTCTTCCTTCTTCTTTTCTTCTTCTTTTCTTCTT-CTTCT 145125

Query:    584 -CTTCTCCTTCTT-TCTTCCTCGTTCCCCTTTCC-CTCCCTCCTCTCTTTCATTCCTTTT 640
               ||| || ||||| | ||| || |||  ||||||  || | |||| |||| | ||  | ||
Sbjct: 145126 TCTTTTC-TTCTTCTTTTCTTC-TTCTTCTTTCTTCTTCTTCCT-TCTT-CTTTT-TCTT 145180

Query:    641 CCTATTTTTTGGCTTCCTCGTTCCGTCCTCCTCTTCCTCTTATTCTCCTCTTTCTTCTTC 700
              | | || |||  |||| || |||| || ||||||||||||| |||  |||  ||||| ||
Sbjct: 145181 CTTCTTCTTTCTCTTCTTC-TTCC-TCTTCCTCTTCCTCTTCTTCCTCTCCCTCTTCCTC 145238

Query:    701 -TCTTGCTCCCCCATCC-CCTCTTC-TCCAT-TACTT-TCTTCTTCGCTCTCCCTATTCT 755
               || | |||| |  ||| ||||||| |||   | ||  || |||||   |||| | |||
Sbjct: 145239 CTCCTCCTCCTCT-TCCTCCTCTTCCTCCTCCTCCTCCTCCTCTTCTTCCTCC-TCTTCC 145296

Query:    756 TGCTCCTCTTCTCTCTGTCCTTCGTC-TCCTCCTCT-CTTTCGCTCCCTCATTGCTTTTC 813
              |  || ||||||  || | |||| || || |||||| ||| | || |||| || || |||
Sbjct: 145297 TCTTCTTCTTCTTCCTCTTCTTCTTCCTCTTCCTCTTCTTCCTCTTCCTC-TTCCTCTTC 145355
```


© 2001 deNovo Biologic | www.denovo-bio.com

```
Query:    814 CCCGCTTCTTCT-CCTCT-CTCTCCTCCTCTA-TTCACGACTTCT-CAGCTCCNCGTTCT 869
                  |||||||| ||||| || || || |||  ||| |  ||||| |  || | | ||||
Sbjct: 145356 TTT-CTTCTTCTTCCTCTTCTTTCTTCTTCTTCTTC-CT-CTTCTTCTTCTTCTCCTTCT 145412

Query:    870 CTCCCTCCTCTCTCTCTCTCCCTCTCTCTCCTCCTCCTCATGCTC-TGCTCCTTCT 928
               ||  || ||| |||  |||| ||||| ||| || |||||||| | ||| | |||||||
Sbjct: 145413 TTCT-TCTTCT-TCTTCCTCTTCCTCT-TCT-TC-TCCTCCTCCTTCTCCTTCTCCTTCC 145467

Query:    929 GTCTCTCTCTTTCA-CTACTTGATTT-CTTAGTCCCTC-TCTTCCGATCTTACCTGTATC 985
               ||| |||| |||  || ||   |   |||  ||| || |||||   | |  |||    |
Sbjct: 145468 -TCT-TCTCCTTCCTCTTCTCCTTCCACTTC-TCCTTCCTCTTCTCCT-TCCCCTTCCCC 145523

Query:    986 TCTTCCTCCATCTGTTCCCCCATATACAATTCTTTGCCCCCGTCTCGCTTCGTCTGTCGC 1045
              || | ||||  ||  ||| || | | |  | | |  || || |||| |||| |||  |
Sbjct: 145524 TCCTTCTCCTCCTCCTCCTCCGTCTCC--TCCGTCTCCTCCATCTC-CTTCATCTC-CTT 145579

Query:   1046 ACTATCGTCTCTCTGTGTTCGACT-CTC-TCCTCCC-CCTCGTTCTCCCGTCC-CT-CCC 1100
               || || ||| ||| | |||  || ||| | ||||  |||| ||||||  ||| || |||
Sbjct: 145580 -CT-TCTTCT-TCT-TCTTCTCCTTCTCCTTCTCCTTCCTC-TTCTCCT-TCCTCTTCCC 145633

Query:   1101 TGACTCTCTCCTCTCTC-TCTGTAACCTCTCATCACCT-CTCCT-C-CATTGCTCTCATC 1156
                 || ||||||| ||| || ||  |||| |||| ||| ||||| | | ||  |||  ||
Sbjct: 145634 CTCCT-TCTCCTC-CTCCTCCGTCTCCTC-CATCTCCTTCTCCTTCTCCTTCTTCTT-TC 145689

Query:   1157 TCTTTCGNAGTCTTCGTTGC-CTGCTATCTTT-TTCTT-TCCCCCTCCC 1202
              || |||    ||||| ||   || || ||||| ||||| | || || ||
Sbjct: 145690 TCCTTCTTCTTCTTCTTTCTTCTTCT-TCTTTCTTCTTCTTCCTCTTCC 145737
```


© 2001 deNovo Biologic | www.denovo-bio.com