Rapid evolution of cancer/testis genes on the X chromosome

Brian J Stevenson*1, Christian Iseli1, Sumir Panji2, Monique Zahn-Zabal1, Winston Hide2, Lloyd J Old3, Andrew J Simpson3 and C Victor Jongeneel1

Address: 1Ludwig Institute for Cancer Research and Swiss Institute of Bioinformatics, CH-1015 Lausanne, Switzerland, 2South African National Bioinformatics Institute, University of the Western Cape, Bellville, 7535, South Africa and 3Ludwig Institute for Cancer Research, New York Branch at Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10021, USA

Email: Brian J Stevenson* - brian.stevenson@licr.org; Christian Iseli - christian.iseli@licr.org; Sumir Panji - sumir@sanbi.ac.za; Monique Zahn-Zabal - monique.zahn@licr.org; Winston Hide - winhide@sanbi.ac.za; Lloyd J Old - lold@licr.org; Andrew J Simpson - asimpson@licr.org; C Victor Jongeneel - victor.jongeneel@licr.org

* Corresponding author

Abstract

Background: Cancer/testis (CT) genes are normally expressed only in germ cells, but can be activated in the cancer state. This unusual property, together with the finding that many CT proteins elicit an antigenic response in cancer patients, has established a role for this class of genes as targets in immunotherapy regimes. Many families of CT genes have been identified in the human genome, but their biological function for the most part remains unclear. While it has been shown that some CT genes are under diversifying selection, this question has not been addressed before for the class as a whole.

Results: To shed more light on this interesting group of genes, we exploited the generation of a draft chimpanzee (Pan troglodytes) genomic sequence to examine CT genes in an organism that is closely related to human, and generated a high-quality, manually curated set of human:chimpanzee CT gene alignments. We find that the chimpanzee genome contains homologues to most of the human CT families, and that the genes are located on the same chromosome and at a similar copy number to those in human. Comparison of putative human:chimpanzee orthologues indicates that CT genes located on chromosome X are diverging faster and are undergoing stronger diversifying selection than those on the autosomes or than a set of control genes on either chromosome X or autosomes.

Conclusion: Given their high level of diversifying selection, we suggest that CT genes are primarily responsible for the observed rapid evolution of protein-coding genes on the X chromosome.
nificant number of additional CT genes [4,5], against most of which immune responses have not yet been doc-
dumented. Nevertheless, all CT genes are in principle attractive targets for cancer immunotherapy, because the gonads are immunoprivilged organs and anti-CT immune responses will therefore target tumours specifically. Vaccination using peptides derived from the NY-
Eso-1 (CTAG1B) and MAGEA1 CT genes has already been proven to bring clinical benefits to melanoma patients [6,7].

CT genes comprise more than 240 members from 70 families, and can be subdivided into two broad categories based on chromosomal localization. CT-X genes are located on the X chromosome, are mostly members of gene families organized into complex direct and inverted repeats, and are expressed primarily during the spermatogonial stage of spermatogenesis [8]. Non-X CT genes are located on autosomes, are mostly single-copy genes, and are expressed primarily during the meiotic and reduction division stages of spermatogenesis [8]. Careful annotation of the sequence of the human X chromosome has revealed that as many as 10% of all genes present on the chromosome are members of known CT families [9]; further analysis of the expression patterns of genes of unknown function located in repeated regions could even increase this estimate [5]. The biological functions of most CT-X genes have not been characterized in any detail. However, evidence is emerging that the best studied of these, the MAGE genes, can act as signal transducing transcriptional modulators. Moreover, MAGE genes appear to be able to mediate proliferative signals [10-12] and a member of the GAGE family has been shown to repress apoptosis [13], thus directly contributing to the malignant phenotype when aberrantly expressed in cancer. Available data suggest that many CT genes are involved in the re-programming of the transcriptional machinery that occurs during the transition from mitotic to meiotic division during spermatogenesis. It has been suggested that a similar re-programming may be responsible for some of the phenotype of malignant cancer cells [8,14].

There is mounting evidence that the evolutionary history of the human X chromosome is significantly different from that of autosomes. It contains a disproportionate number of tandem and interspersed segmental duplications, both direct and inverted, containing genes with a testis-specific expression pattern including many CT-X genes [9]. These duplications are unstable in the genome, and subject to copy number polymorphisms, both within the human population and between humans and chim-
pancees [15,16]. While its overall DNA sequence has diverged significantly less than that of autosomes since speciation of hominoids from chimpanzees [17], a signif-
icant proportion of protein-coding genes located on the X chromosome are under higher diversifying (positive) selection than those on autosomes [18]. Genes located on the X chromosome are also the most abundant source of functional retrogenes in the primate lineage, and constitute a reservoir of genetic material for the generation of new genes and functions in this lineage, again with a bias toward testis-specific functions [19,20].

For all of these reasons, it is of interest to trace the evolution history of CT genes, and particularly of the CT-X subset, and to measure the selective pressures that act on them. Many of the human CT-X genes do not have easily identifiable orthologues in the mouse, rat or dog genomes, precluding such an analysis among Eutheria using currently available genome data. For example, it has been shown that the large MAGE family of CT-X genes has expanded independently in the primate and rodent lineages [21]. The recent availability of a draft genome for the chimpanzee has made it feasible to study the evolution of the CT genes within the primate lineage. We show here that the CT genes in general and the CT-X genes in partic-
ular are under strong diversifying pressure and amongst the fastest-evolving genes in the human genome.

Results

Identification of CT gene families in chimpanzee
To date at least seventy CT gene families, many with multiple members, have been identified in human. We took the opportunity afforded by the publication of the initial sequence of the chimpanzee genome [18] to ask whether CT genes were conserved in man's closest evolutionary neighbour. To this end we assembled a list of human tran-
script sequences representing all CT gene families, and searched for homologous sequences in the human and chimpanzee genomes. We expected that given the relatively short time elapsed since human-chimp divergence (~ 6 million years ago [17]) the human sequences would be able to detect CT gene homologues in the chim-
panzee genome. Moreover, since the majority of CT genes isolated thus far were detected and characterized using transcript information via cDNA cloning protocols, per-
forming the same search in human allowed us to identify all CT genes present in the current assembly of the human genome. We implemented a two-stage approach in order to accurately define the structure of each CT gene locus. First, we used MegaBlast [22] to search for regions homol-
ogous to the CT transcript sequences. Then we applied the SiBSim4 cDNA to genome alignment program (an improved version of sim4 [23]) to these regions to estab-
lish a gene structure from a locus-specific spliced align-
ment (see Methods). As can be seen in Table 1, almost all human CT families are found in chimpanzee, and the chromosomal locations of the CT genes in chimpanzee correspond to those in human. In terms of copy number, the biggest family, PRAME, is well represented in chim-
Table 1: Number and chromosomal location of CT genes in human and chimpanzee

| CT Number | Family Name | Human Chromosome | Human Gene Number | Chimpanzee Chromosome | Chimpanzee Gene Number |
|-----------|-------------|------------------|-------------------|-----------------------|------------------------|
| CT1       | MAGEA       | X                | 13 (0)            | X                     | 9 (0)                  |
| CT2       | BAGE        | 5, 7, 9, 18, 21  | 7 (0)             | 7, 9, 18              | 4 (0)                  |
| CT3       | MAGEB       | X                | 7 (1)             | X                     | 7 (1)                  |
| CT4       | GAGE        | X                | 16 (0)            | X                     | 3 (0)                  |
| CT5       | SSX         | X                | 14 (0)            | X                     | 8 (0)                  |
| CT6       | CTAG        | X                | 3 (0)             | X                     | 1 (0)                  |
| CT7       | MAGEC       | X                | 2 (0)             | X                     | 1 (0)                  |
| CT8       | SYCP1       | 1                | 1 (0)             | 1                     | 1 (0)                  |
| CT9       | BRDT        | 1                | 1 (0)             | 1                     | 1 (0)                  |
| CT10      | MAGEE       | X                | 2 (2)             | X                     | 1 (1)                  |
| CT11      | SPANX       | X                | 11 (0)            | X                     | 4 (0)                  |
| CT12      | XAGE        | X                | 14 (0)            | 0                     | 12 (0)                 |
| CT13      | DDX43       | 6                | 1 (0)             | 6                     | 1 (0)                  |
| CT14      | SAGE        | X                | 1 (0)             | X                     | 1 (0)                  |
| CT15      | ADAM2       | 4, 8             | 2 (0)             | 4, 8                  | 2 (0)                  |
| CT16      | PAGE        | X                | 7 (0)             | X                     | 6 (0)                  |
| CT17      | LIPI        | 21               | 2 (0)             | -                     | 0 (0)                  |
| CT21      | CTAGE       | 2, 6, 7, 9, 10, 13, 14, 18 | 21 (12) | 2B, 6, 7, 9, 10, 13, 14, 18 | 15 (6) |
| CT24      | CSAG        | X                | 4 (0)             | X                     | 2 (0)                  |
| CT25      | DSCR8       | 21               | 2 (0)             | -                     | 0 (0)                  |
| CT26      | DDX53       | X                | 1 (1)             | X                     | 1 (1)                  |
| CT27      | CTCLF       | 20               | 1 (0)             | 20                    | 1 (0)                  |
| CT28      | LUZP4       | X                | 1 (0)             | X                     | 1 (0)                  |
| CT29      | CASC5       | 15               | 1 (0)             | 15                    | 1 (0)                  |
| CT30      | TDP53       | 13, 15, X        | 4 (3)             | 15, X                 | 2 (2)                  |
| CT32      | LDHC        | 11               | 1 (0)             | 11                    | 1 (0)                  |
| CT33      | MORCI       | 3                | 1 (0)             | 3                     | 1 (0)                  |
| CT34      | DKKL1       | 19, 20           | 2 (1)             | 19, 20                | 2 (1)                  |
| CT35      | SPO11       | 20               | 1 (0)             | 20                    | 1 (0)                  |
| CT36      | CRISP2      | 6                | 1 (0)             | 6                     | 1 (0)                  |
| CT37      | FMRI1NB     | X                | 1 (0)             | X                     | 1 (0)                  |
| CT38      | FTHL17      | X                | 4 (4)             | X                     | 5 (5)                  |
| CT39      | NXF2        | X                | 2 (0)             | X                     | 1 (0)                  |
| CT41      | TDRD        | 6, 10            | 2 (0)             | 6, 10                 | 2 (0)                  |
| CT42      | TEX15       | 8                | 1 (0)             | 8                     | 1 (0)                  |
| CT43      | FATE1       | X                | 1 (0)             | X                     | 1 (0)                  |
| CT44      | TPTE        | 13, 21, Y        | 4 (0)             | 13                    | 1 (0)                  |
| CT45      | CT45        | X                | 6 (0)             | X                     | 4 (0)                  |
| CT46      | HORMAD1     | 1, 6             | 2 (1)             | 1, 6                  | 2 (1)                  |
| CT47      | LOC2535313  | X                | 12 (0)            | X                     | 2 (0)                  |
| CT48      | SLCO6A1     | 5                | 1 (0)             | 5                     | 1 (0)                  |
| CT49      | TAG         | 5                | 1 (0)             | 5                     | 1 (0)                  |
| CT50      | LEMDI       | 1                | 1 (0)             | 1                     | 1 (0)                  |
| CT51      | HSPB9       | 17               | 1 (1)             | 17                    | 1 (1)                  |
| CT53      | ZNF165      | 6                | 1 (0)             | 6                     | 1 (0)                  |
| CT54      | SPACA3      | 17               | 1 (0)             | -                     | 0 (0)                  |
| CT55      | CXor48      | X                | 3 (0)             | X                     | 1 (0)                  |
| CT56      | THEG        | 19               | 1 (0)             | 19                    | 1 (0)                  |
| CT57      | ACTL8       | 1                | 1 (0)             | 1                     | 1 (0)                  |
| CT58      | NALP4       | 19               | 1 (0)             | 19                    | 1 (0)                  |
| CT59      | COX6B2      | 19               | 1 (0)             | 19                    | 1 (0)                  |
| CT60      | BC047459    | 15               | 2 (0)             | Un                    | 1 (0)                  |
| CT61      | CCD33       | 15               | 1 (0)             | 15                    | 1 (0)                  |
| CT62      | BC048128    | 15               | 1 (0)             | 15                    | 1 (0)                  |
| CT63      | PASD1       | X                | 1 (0)             | X                     | 1 (0)                  |
| CT65      | TULP2       | 19               | 1 (0)             | 19                    | 1 (0)                  |
| CT66      | AA884595    | 7                | 1 (1)             | 7                     | 1 (1)                  |
| CT68      | MGC27016    | 4                | 1 (0)             | 4                     | 1 (0)                  |
| CT69      | BC040308    | 6                | 1 (0)             | 6                     | 1 (0)                  |
and chimpanzee ORF sequences (see below). Given the
this by analysing the substitution rates between human
between the CT group and the control group. We tested
logues (97.8% versus 98.7%; p < 2.2e-16 by a chi-squared
were on average less closely related than the control ortho-
finding that the chimpanzee and human CT orthologues
tables (99.6% and 98.7%, respectively. The
average identity of the human and
alignments of the same quality as for the CT genes, and
fold: first, this allowed us to generate manually curated
somal chromosomes 18 and 19 (see Methods). The rea-
group, we also defined a group of human-chimpanzee
were interested in the characteristics of CT genes as a
sequences was 99.6% and 97.8%, respectively. Since we
chimpanzee orthologues to the human tran-
55x123]ual sequences was 99.6% and 98.7%, respectively. The rate of synonymous changes (dS) indicates
for the CT genes and controls by chromosomal location.
We estimated the divergence rates of the CT genes from
pairwise sequence alignments of the human and chimpan-
are using phylogenetic analysis (PAML
package [24]). Mutations in a protein-coding gene can either have no effect (synonymous changes) or alter the
sequence of the encoded protein (non-synonymous changes). The rate of synonymous changes (dS) indicates
the type of evolutionary pressure acting on the gene. A dN/dS ratio value less than 1 suggests negative or
function 1). The average identity of the human and
 chimpanzee orthologues to the human transcript
sequence. Ninety-eight orthologous CT
genes, as are MAGEA (9 genes) CTAGE (15
genes), XAGE (12 genes) and SSX (8 genes). The number of
CT genes in each family is probably underestimated
because of the relatively low sequence coverage in the cur-
rent version of the chimpanzee genome assembly. This is
especially true for the X chromosome, where the sequence
coverage is only about 2-fold [18], and where most of the
human multi-gene CT families are located. Nevertheless,
the current data indicate that some chimpanzee CT fami-
lies (FTHL17/CT38, TSPY/CT78 and PRAME) may con-
tain more members than in human.

In order to investigate more closely the relatedness of CT
genes in these two species, we sought putative human and chimpanzee orthologues for as many CT genes as possible,
based on nucleotide sequence identity to the cognate
human transcript sequence. Ninety-eight orthologous CT
pairs were defined in this way (see Methods and addi-
tional file 1). The average identity of the human and
chimpanzee orthologues to the human transcript
sequences was 99.6% and 97.8%, respectively. Since we
were interested in the characteristics of CT genes as a
group, we also defined a group of human-chimpanzee
orthologous non-CT control genes from chromosome X,
where most of the CT genes are located, and from auto-
somal chromosomes 18 and 19 (see Methods). The rea-
sons for choosing a limited set of control genes were two-
fold: first, this allowed us to generate manually curated
alignments of the same quality as for the CT genes, and
second, it provided test and control groups of similar sizes
for statistical analysis. The average identity of the human
and chimpanzee control orthologues to the human tran-
script sequences was 99.6% and 98.7%, respectively. The
finding that the chimpanzee and human CT orthologues
were on average less closely related than the control ortho-
lologues (97.8% versus 98.7%; p < 2.2e-16 by a chi-squared
test) suggested a possible difference in the divergence rates
between the CT group and the control group. We tested
this by analysing the substitution rates between human
and chimpanzee ORF sequences (see below). Given the
high accuracy of the human genomic sequence, the find-
ing that the average human identity was less than 100%
for both CT genes and non-CT control genes presumably
reflects polymorphisms and/or sequencing errors in the
original transcript sequences.

**CT genes on chromosome X are evolving faster than those on other chromosomes**

We estimated the divergence rates of the CT genes from
pairwise sequence alignments of the human and chimpan-
zee orthologues using phylogenetic analysis (PAML
package [24]). Mutations in a protein-coding gene can either have no effect (synonymous changes) or alter the
sequence of the encoded protein (non-synonymous changes). The rate of synonymous changes (dS) indicates
the background mutation frequency, while the ratio of the
non-synonymous to synonymous mutation rates (dN/dS)
dicates the type of evolutionary pressure acting on the
gene. A dN/dS ratio value less than 1 suggests negative or
purifying selection, a ratio equal to 1 suggests neutral evo-
uletion, and a ratio greater than 1 suggests positive or diver-
sifying selection [25]. To test what type of evolutionary
pressure might be acting on the CT genes, we aligned the
ORFs in the human-chimpanzee orthologue pairs and
used the codeml program from the PAML package [24] to
estimate the dN/dS ratios. Again, for comparison pur-
poses, the control genes were subjected to an identical
procedure. Figure 1 shows the distribution of dN/dS ratios
for the CT genes and controls by chromosomal location.
In contrast to the control genes, which show the distribu-
tion of ratios expected if most genes are under purifying
selection, CT genes located on chromosome X have an
excess of ratios greater than one. At the level of individual
genes, SSX1, PAGE2B, SSX4, MAGEB2, GAGE4 and
CPXCR1 have rate ratios greater than 2, indicative of
strong evolutionary selective pressure acting on the gene
products (Table 2). CT genes located on chromosomes
other than chromosome X (CT-nonX) have a distribution of
ratios skewed towards lower values, suggesting that this
subgroup is evolving slower than the CT-X genes. In con-

| CT gene families | Number of genes | Chromosomal location | Average identity | Number of intronless gene copies |
|------------------|-----------------|----------------------|-----------------|-------------------------------|
| CTNA PRAME       | 1, 22           | X                    | 97.8%           | 0 (0)                         |
| CT74 TSPY1       | Y               |                      | 97.8%           | 22 (0)                        |
| CT75 BC03986     | 2               |                      | 97.8%           | 2B (0)                        |
| CT76 SYCEI       | 10              | X                    | 97.8%           | 10 (0)                        |
| CT77 CPXCR1      | X               |                      | 97.8%           | 1 (0)                         |
| CT80 ARMC3       | 2, 21           |                      | 97.8%           | 2A (0)                        |
| CT81 TSGA        | 10              |                      | 97.8%           | 10 (0)                        |
| CT71 SPINLW1     | 2               |                      | 97.8%           | 0 (0)                         |
| CT72 TSSK6       | 19              |                      | 97.8%           | 0 (0)                         |
| CT73 ADAM29      | 4               |                      | 97.8%           | 1 (0)                         |
| CT74 CCDC36      | 3               |                      | 97.8%           | 3 (0)                         |
| CT77 SYCEI       | 10              |                      | 97.8%           | 10 (0)                        |
| CT78 CPXCR1      | X               |                      | 97.8%           | 1 (0)                         |
| CT80 TSGA        | 2, 21           |                      | 97.8%           | 2A (0)                        |
| CT81 TSSK6       | 19              |                      | 97.8%           | 0 (0)                         |

CT gene families are presented in numerical order according to proposed nomenclature [1]. The largest family, PRAME, has not yet been assigned official CT designation. Total gene number for each family was determined according to sequence identity and completeness (see Methods). Numbers in brackets denote the number of intronless gene copies, which in the case of multi-exon genes may indicate putative retrocopy genes.
Table 2: Nucleotide substitution rates estimated from alignments of human and chimpanzee orthologous CT ORFs

| Gene Name     | Refseq  | Chromosome | dN   | dS   | dN/dS  |
|---------------|---------|------------|------|------|--------|
| ACTL8         | NM_030812 | 1          | 0.0012 | 0.0170 | 0.0700 |
| BRTD          | NM_207189 | 1          | 0.0066 | 0.0071 | 0.9216 |
| HORMAD1       | NM_032132 | 1          | 0.0680 | 0.0104 | 0.6500 |
| LEMD1         | NM_001001552 | 1      | 0.0044 | 0.0327 | 0.1342 |
| PRAMEF1       | NM_023013 | 1          | 0.0162 | 0.0288 | 0.5714 |
| PRAMEF2       | NM_023014 | 1          | 0.0304 | 0.0317 | 0.9573 |
| PRAMEF3       | NM_001013692 | 1     | 0.0223 | 0.0269 | 0.8278 |
| PRAMEF4       | NM_001009611 | 1    | 0.0284 | 0.0305 | 0.9314 |
| PRAMEF5       | NM_001013407 | 1    | 0.0353 | 0.0586 | 0.6025 |
| PRAMEF6       | NM_001010889 | 1    | 0.0142 | 0.0149 | 0.9479 |
| PRAMEF8       | NM_001012276 | 1    | 0.0141 | 0.0262 | 0.5383 |
| PRAMEF9       | NM_001039361 | 1    | 0.0184 | 0.0262 | 0.7029 |
| SYCP1         | NM_003176  | 1          | 0.0050 | 0.0123 | 0.4093 |
| BX103208      | BX103208  | 3          | 0.0000 | 0.0346 | 0.0009 |
| CCDC36        | NM_178173 | 3          | 0.0065 | 0.0118 | 0.5502 |
| MORC1         | NM_014429 | 3          | 0.0071 | 0.0112 | 0.6325 |
| CDC110        | NM_152775 | 4          | 0.0081 | 0.0142 | 0.5694 |
| MGC27016      | NM_144979 | 4          | 0.0017 | 0.0166 | 0.0994 |
| SLC06A1       | NM_173488 | 5          | 0.0083 | 0.0093 | 0.8940 |
| TAG1          | AY328030  | 5          | 0.0001 | 0.1321 | 0.0009 |
| BC040308      | BC040308  | 6          | 0.0381 | 0.0004 | 0.∞    |
| CNIP2         | NM_003296 | 6          | 0.0034 | 0.0078 | 0.4355 |
| DDX43         | NM_018665 | 6          | 0.0046 | 0.0084 | 0.5422 |
| TDRD6         | NM_001010870 | 6   | 0.0029 | 0.0077 | 0.3756 |
| ZNF165        | NM_003447 | 6          | 0.0028 | 0.0083 | 0.3332 |
| AAB84595      | AA884595  | 7          | 0.0000 | 0.0000 | 0.4503 |
| BAGE2         | NM_182482 | 7          | 0.0000 | 0.0000 | 0.4741 |
| ADAM2         | NM_001464 | 8          | 0.0090 | 0.0102 | 0.8787 |
| TEX15         | NM_031271 | 8          | 0.0064 | 0.0103 | 0.6188 |
| BAGE          | NM_001187 | 9          | 0.0000 | 0.0441 | 0.0009 |
| ARM3C         | NM_173081 | 10         | 0.0049 | 0.0142 | 0.3479 |
| SYCE1         | NM_130784 | 10         | 0.0073 | 0.0105 | 0.6979 |
| TDRD1         | NM_198795 | 10         | 0.0035 | 0.0085 | 0.4101 |
| LDHC          | NM_002301 | 11         | 0.0000 | 0.0070 | 0.0009 |
| TPTE          | NM_199261 | 13         | 0.0118 | 0.0095 | 1.2398 |
| CTAGE5        | NM_203356 | 14         | 0.0029 | 0.0082 | 0.3578 |
| BCO48128      | BCO48128  | 15         | 0.0077 | 0.0143 | 0.5355 |
| CASC5         | NM_170589 | 15         | 0.0084 | 0.0116 | 0.7226 |
| CDC23         | NM_182791 | 15         | 0.0093 | 0.0192 | 0.4835 |
| Kibb4         | XM_375358 | 16         | 0.0051 | 0.0109 | 0.4713 |
| HSPB9         | NM_033194 | 17         | 0.0112 | 0.0184 | 0.6077 |
| CTAGE1        | NM_172241 | 18         | 0.0108 | 0.0204 | 0.5311 |
| COX6B2        | NM_144613 | 19         | 0.0047 | 0.0138 | 0.3413 |
| DKKL1         | NM_014419 | 19         | 0.0055 | 0.0060 | 0.9034 |
| NALP4         | NM_134444 | 19         | 0.0090 | 0.0180 | 0.4981 |
| THEG          | NM_016585 | 19         | 0.0100 | 0.0091 | 1.1002 |
| TULP2         | NM_003323 | 19         | 0.0059 | 0.0056 | 1.0501 |
| CTFL          | NM_0080618 | 20       | 0.0124 | 0.0169 | 0.7316 |
| SPINELW1      | NM_181502 | 20         | 0.0134 | 0.0262 | 0.5122 |
| SPO11         | NM_01244 | 20         | 0.0044 | 0.0119 | 0.3679 |
| PRAME         | NM_006115 | 22         | 0.0191 | 0.0162 | 1.1798 |
| CPXCR1        | NM_033048 | X          | 0.0104 | 0.0047 | 2.3411 |
| CSAG1         | NM_153478 | X          | 0.0622 | 0.0006 | 0.∞    |
| CSAG2         | NM_004909 | X          | 0.0163 | 0.0266 | 0.6138 |
| CT45-2        | NM_152582 | X          | 0.0207 | 0.0002 | 0.∞    |
| DDX53         | NM_182699 | X          | 0.0159 | 0.0109 | 1.4567 |
| FATE1         | NM_033085 | X          | 0.0025 | 0.0142 | 0.1755 |
| FMRIINB       | NM_152578 | X          | 0.0374 | 0.0228 | 1.6405 |
The majority of control genes, irrespective of chromosomal location, have rate ratios less than 0.5, suggestive of purifying selection. In addition, the nonsynonymous substitution rates for CT genes which had no synonymous changes between human and chimpanzee was on average higher than for the controls (see additional file 2).

The apparent difference between the dN/dS distributions for the CT genes and the controls was assessed for significance using a nonparametric Mann-Whitney test, which indicates whether the medians of the two populations are significantly different. The difference in dN/dS values between all CT genes and all controls is highly significant with a p-value of 1.128e-11 (Table 3). Moreover, the difference between CT genes and the controls is significant whether the CT genes are located on chromosome X (p = 4.686e-10) or not (p = 1.498e-05). The distribution of dN/dS values is also significantly different for CT genes on chromosome X compared to those elsewhere (p = 2.812e-05), suggesting that there is stronger selective pressure on CT genes located on chromosome X. In contrast, there is no significant difference in the distribution of dN/dS ratios between the control genes located on chromosome X or elsewhere (p = 0.4962). Previous work has shown that the protein-coding genes on the hominid X chromosome have a higher average dN/dS value than other chromosomes [18]. Our results suggest that the CT genes contribute strongly to this difference, and thus to the rapid evolution of protein-coding genes on the X chromosome.

**Discussion**

Several recent publications have taken advantage of the chimpanzee draft genome to identify genes that are under diversifying selection in the primate lineage ([26] and ref-
The proportion of genes in each category with ratios in intervals A-I is shown. The categories are: CT-X, CT genes on chromosome X (N = 33); CT-nonX, CT genes not on chromosome X (N = 64); Control-X, control genes on chromosome X (N = 49); Control-nonX, control genes not on chromosome X (N = 33); CT-nonX, CT genes not on chromosome X (N = 71). The intervals are: \(0 \leq A < 0.25; 0.25 \leq B \leq 0.5; 0.5 < C \leq 0.75; 0.75 < D \leq 1.0; 1.0 < E \leq 1.25; 1.25 < F \leq 1.5; 1.5 < G \leq 1.75; 1.75 < H \leq 2.0; 2 < I \leq 4.0\). Genes which had no synonymous changes (dN/dS denoted '∞' in Table 2) were omitted from this analysis. For comparisons, differences in the distributions were also assessed for significance using a parametric Welch two sample t-test; see additional file 3.

Table 3: Significance of the differences in the distributions of dN/dS ratios between CT and control ORFs

| Comparison                          | p-value  |
|-------------------------------------|----------|
| All CTs vs. All controls            | 6.22e-12 |
| CT-Xs vs. Control-Xs                | 2.31e-10 |
| Non-X CTs vs. Non-X controls        | 1.50e-05 |
| CT-Xs vs. Non-X CTs                 | 1.62e-05 |
| Controls on X vs. Non-X controls    | 0.50     |

The distributions of dN/dS ratios from groups of CT and control ORFs were compared with each other, and any difference assessed using the non-parametric Mann-Whitney rank sum test [43]. Ratios denoted by '∞' in Table 2 were omitted from this analysis. For comparison, differences in the distributions were also assessed for significance using a parametric Welch two sample t-test; see additional file 3.

Figure 1
Distribution of dN/dS ratios for CT genes and controls. The proportion of genes in each category with ratios in intervals A-I is shown. The categories are: CT-X, CT genes on chromosome X (N = 33); CT-nonX, CT genes not on chromosome X (N = 64); Control-X, control genes on chromosome X (N = 49); Control-nonX, control genes not on chromosome X (N = 33); CT-nonX, CT genes not on chromosome X (N = 71). The intervals are: \(0 \leq A < 0.25; 0.25 \leq B \leq 0.5; 0.5 < C \leq 0.75; 0.75 < D \leq 1.0; 1.0 < E \leq 1.25; 1.25 < F \leq 1.5; 1.5 < G \leq 1.75; 1.75 < H \leq 2.0; 2 < I \leq 4.0\). Genes which had no synonymous changes (dN/dS denoted '∞' in Table 2) were omitted from this analysis. For comparisons, differences in the distributions were also assessed for significance using a parametric Welch two sample t-test; see additional file 3.

Table 4: Reports of positive selection pressure on CT genes

| CT family | Gene name | Human RefSeq | Reference | Present work# |
|-----------|-----------|--------------|-----------|---------------|
| CT1       | MAGEA4    | NM_002362    | I         | Yes           |
| CT1       | MAGEA5    | NM_021049    | I         | Yes           |
| CT1       | MAGEA10   | NM_021048    | I         | Yes           |
| CT2       | BAGE2     | NM_182482    | I         |               |
| CT3       | MAGEB2    | NM_002364    | I         | Yes           |
| CT3       | MAGEB3    | NM_002365    | I         | Yes           |
| CT5       | SXX1      | NM_005635    | I, III    | Yes           |
| CT5       | SXX8      | NM_174961    | I, III    | Yes           |
| CT7       | MAGEC2    | NM_016249    | I         |               |
| CT7       | MAGEC3    | NM_138702    | I         |               |
| CT11      | SPANX-N2  | NM_00109615  | III       |               |
| CT11      | SPANX-N3  | NM_00109609  | III       |               |
| CT11      | SPANX-N4  | NM_00109613  | III       |               |
| CT11      | SPANX-N5  | NM_00109616  | III       |               |
| CT11      | SPANX8    | NM_013453    | III       |               |
| CT11      | SPANXC    | NM_022661    | III       |               |
| CT14      | SAGE1     | NM_018666    | I, II     | Yes           |
| CT16      | PAGE1     | NM_003785    | I         | Yes           |
| CT37      | FMR1NB    | NM_152578    | I         | Yes           |
| CT38      | FTHL17    | NM_031894    | I         | Yes           |
| CT48      | SLCO6A1   | NM_173488    | I         |               |
| CT55      | CXXorf48  | NM_017863    | I         |               |
| CT56      | THEG      | NM_016585    | I         | Yes           |
| CT63      | PASD1     | NM_173493    | I         |               |
| CT65      | TULP2     | NM_003323    | I         | Yes           |
| CT77      | CPXCR1    | NM_033048    | I         | Yes           |
| CT80      | PIWIL2    | NM_018068    | I         |               |
| CTNA      | PRAME     | NM_006115    | I         | Yes           |
| CTNA      | PRAME     | cluster on chromosome I | IV | Yes |

Positive selection pressure on CT genes, from analysis of human and chimpanzee sequences, reported in: I, as defined by dN/dS > 1 [18, 33]. II, as defined by likelihood ratio test with p-value < 0.05 [35]. III, as defined by dN/dS > 1 [27]. IV, inferred from dN/dS > 1 and sites modelling on human alignments [28]. # Confirmed 16 previously reported positively selected CT genes, plus an additional 18 positively selected CT genes (see Table 2).
of CT homologues present in the chimpanzee genome, and some of the human:chimpanzee pairs may not correspond to true orthologues. However, neither of these problems should significantly affect the main conclusions of our study.

Given the close evolutionary kinship between humans and chimpanzees it is not surprising that all known CT gene families are shared between the two species. On the other hand, homologues of many CT antigens have not been found outside the primate lineage so far, and the available genome data are still too sparse to track the appearance of CT gene families during mammalian evolution. Even though the data are still incomplete, it is clear that most CT gene families are undergoing copy number expansions in the primate lineage, presumably driven by non-allelic homologous recombination between segmental duplications. The best-studied CT family in this respect is SPANX, which is present as a single-copy gene in rodents and has duplicated and acquired new sub-families in the primate lineage, including at least one (SPANX-C) found to be specific to humans on the basis of its genomic position [27]. SPANX genes have been shown to have copy number polymorphisms in the human population, potentially linked to susceptibility to prostate cancer, and to undergo very rapid evolution affecting both dN and dS [29]. An elegant study of the PRAME cluster on human chromosome 1 [28] revealed the recent expansion in the human lineage of these genes via two large segmental duplications, and subsequent smaller duplications that may be polymorphic in the human population. The large MAGE family of CT antigens, which also comprises genes that do not show a CT expression pattern, has expanded in both the primate and rodent lineages, but independently [21]. Our data also show that many MAGE genes are under diversifying selection (Table 2).

By definition, CT genes are expressed in testis, and for those for which data exists expression has been shown to be restricted to cells involved in spermatogenesis. It is believed that many CT genes are also expressed during oogenesis, but data on this process are still very sparse [30,31]. There is abundant evidence in the literature that many genes expressed predominantly during gametogenesis, as well as those implicated in reproduction in general (e.g. those encoding proteins found in the seminal fluid or expressed predominantly in the prostate) are undergoing positive selection during evolution [32-34]. In this respect, CT genes seem to behave much like other reproductive genes.

However, the CT-X genes are a special case, in that diversifying selective pressure seems more intense on this class. It is probable that the evolutionary pressures driving changes in the encoded protein sequences and those driving the expansion of the CT-X gene families are similar. Strikingly, the X chromosome is enriched in intrachromosomal tandem segmental duplications relative to autosomes [9]. Several hypotheses have been put forward to explain why a subset of genes located on the X chromosome is evolving faster than those on autosomes [34-36]. Our data do not shed new light on this subject. However, it is interesting to note that CT-X genes contribute very significantly to the high average positive selection observed in protein-encoding genes on this chromosome, against a genomic background that is much more highly conserved than on the autosomes [17]. One may speculate that transcriptional controls on recently duplicated genes could be relaxed relative to the parental copies, thereby allowing re-expression in tumours and the partial replication in these tumours of the transcriptional changes accompanying gametogenesis.

Conclusions
Essentially all human CT families have homologues at the same chromosomal locations in the chimpanzee genome. The copy numbers in the multi-gene CT families may differ between the two species but until a high-quality assembly of the chimpanzee genome is available this cannot be assessed in a reliable way. On the average, CT genes are under stronger positive selection than a set of randomly selected control genes. CT-X genes as a group are evolving very rapidly, not only relative to control genes on the X chromosome or on autosomes, but also relative to autosomal CT genes.

Methods
CT genes and human/chimpanzee genomic sequences
Human Reference sequence (RefSeq [37]), or GenBank (where no RefSeq was available) entries were obtained for transcripts representing all documented CT gene families in the CT Gene Database [38]. Transcript sequences were also obtained for additional candidate CT genes described in recent publications, which have not yet been added to the CT Gene Database. In some cases, multiple alternatively spliced transcript sequences from the same gene were selected to maximize sequence representation of the locus. Although PRAME has not been designated a CT gene, due to its trace level of expression in some normal adult tissues other than testis, it does exhibit the other main characteristics of CT genes, i.e. strong expression in the testis and up-regulation in various tumours, and was included in the set of CT genes selected for this study. Non-CT control genes were randomly chosen from lists of genes having a RefSeq identifier on chromosomes X, 18 (low gene density) and 19 (high gene density), generated using BioMart [39,40]. Control genes were selected from locations distributed uniformly along the lengths of the chromosomes to average out site-specific differences in mutation rates. The human (Homo sapiens) genomic
sequence used was NCBI Build Number 36 (version 1, release date 9 March 2006), obtained from the NCBI. The chimpanzee (Pan troglodytes) genomic sequence used was NCBI Build Number 2 (version 1, release date 4 October 2006), also obtained from the NCBI.

**Identification of CT gene loci in human and chimpanzee**

CT gene loci were identified in both human and chimpanzee based on sequence identity between the human transcript sequences and human or chimpanzee genomic sequences. We used MegaBlast [22] to identify genomic regions homologous to the RefSeq sequences and SIBsim4 [41] (an improved version of sim4 [23]) to produce high quality spliced alignments at those sites, from which locus-specific transcript sequences were generated. A gene was considered complete if the alignment contained at least 80% of the cognate transcript length or 80% of the annotated open reading frame (ORF), and had at least 85% identity to the human transcript sequence. Putative orthologues were identified as the sequences in human and chimpanzee genomes having the highest identity (and satisfying the 80% length threshold) to the same human transcript sequence. In many cases the poor quality (gaps, incorrect assembly) of the published chimpanzee genome sequence prevented us from finding a chimpanzee orthologue to the human gene. High quality sequence alignments for putative human/chimpanzee orthologues were obtained for 98 of the initial list of 135 CT genes (73%) and 153 of the 180 control genes (85%) selected randomly from chromosomes 18, 19 and X.

**Divergence of CT genes**

The genome-based transcript sequences derived from human and chimpanzee for each putative orthologous pair were aligned using clustalw (version 1.81 [42]), with gap extension penalties set to zero to allow gaps in the alignment arising from sequences missing in the chimpanzee assembly. Both sequences in the alignment were then trimmed to the extent of the human ORF based on annotation in the RefSeq or GenBank entry. Each nucleotide alignment was manually curated and revised, if necessary, to reflect the corresponding protein alignment. ORFs containing stop codons were dropped from the analysis. Rates of synonymous (dS; also known as Ks) and non-synonymous (dN; also known as Ka) substitutions between aligned ORFs were estimated using the codeml programme from the PAML package [24] with the F3x4 codon frequency model (and runmode = -2 in the codeml control file). Note that incomplete codons in either the human or the chimpanzee sequence are ignored by codeml. The statistical significance of differences in the distributions between human-chimpanzee divergence rates (dN/dS) among CT genes and controls was assessed using a Mann-Whitney (Table 3) or Welch two sample t-test (additional file 3) in the R package [43].

**Abbreviations**

CT – cancer/testis

CT-X – CT genes on chromosome X

dN – nonsynonymous substitution rate

dS – synonymous substitution rate

NCBI – National Center for Biotechnology Information

ORF – open reading frame

PAML – phylogenetic analysis by maximum likelihood

**Authors’ contributions**

BJS, CI, LJO, AJS and CVJ designed the experiments. BJS wrote the software pipeline to identify human and chimpanzee CT genes and to produce ORF alignments. SP, MZ and WH scanned the literature for citations of positive selection. BJS and CVJ wrote the manuscript, which was read and approved by all authors.

**Additional material**

**Additional File 1**

Homology data on the human:chimpanzee putative orthologues used in this study. Excel spreadsheet presenting homology data on the human:chimpanzee putative orthologues.

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2164-8-129-S1.xls](http://www.biomedcentral.com/content/supplementary/1471-2164-8-129-S1.xls)

**Additional File 2**

Phylogenetic analysis of CT and control gene ORFs using codeml. Excel spreadsheet presenting data additional to that displayed in Table 2.

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2164-8-129-S2.xls](http://www.biomedcentral.com/content/supplementary/1471-2164-8-129-S2.xls)

**Additional File 3**

Significance of the differences in the distributions of dN/dS ratios between CT and control ORFs using a parametric t-test. Distribution of dN/dS ratios assessed by parametric t-test. The results are qualitatively similar to those presented in Table 3 and confirm that the distribution of dN/dS values is different between CT genes and controls.

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2164-8-129-S3.doc](http://www.biomedcentral.com/content/supplementary/1471-2164-8-129-S3.doc)

**Acknowledgements**

We thank members of SIB Lausanne for discussions, and in particular Asa Wirapati and Frédéric Schutz for advice on statistical analysis of the phylogenetic data. This work was supported by the Ludwig Institute for Cancer Research.
References

1. Scanlan MJ, Simpson AJ, Old LJ: The cancer/testis genes: review, standardization, and commentary. Cancer Immun 2004, 4:1.

2. van der Bruggen P, Traversari C, Chomez P, Lorquin C, De Plaen E, Van den Eynde B, Knuth A, Boon T: A gene encoding an antigen recognized by cytolytic T lymphocytes on a melanoma. Science 1991, 254(5038):1643-1647.

3. Chen YT, Scanlan MJ, Venditti CA, Jongeneel CV, Old LJ, Wyllie AH: A testicular antigen aberrantly expressed in human cancers detected by antibody screening. Proceedings of the National Academy of Sciences of the United States of America 1997, 94(5):1914-1918.

4. Chen YT, Scanlan MJ, Venditti CA, Chua R, Theiler G, Stevenson BJ, Faulkner L, Francis F, Frankland J, Fraser AE, Galgoczy P, Gilbert J, Gill I, Eades T, Ellwood M, Emery-Cohen A, Errington H, Evans KL, Hurles ME, Andrews TD, Scott CE, Searle S, Ramser J, Whittaker A, Howe KL, Ashurst JL, Fulton RS, Sudbrak R, Wen G, Jones MC, Platzer M, Howell GR, Burrows C, Bird CP, Frankish A, Lovell FL, DM, Loulseged H, Loveland JE, Lovell JD, Lozado R, Lu J, Lyne R, Ma GK, Langford C, Lawlor S, Leversha M, Lewis L, Liu W, Lloyd C, Lloyd D, Louisegees L, Lovell J, Lu J, Lynne R, Ma J, Maheshwari M, Matthews LH, McDowell J, McLaren S, McMurray A, Meid P, Meitinger T, Milne S, Miner G, Mistley SL, Morgan M, Morris S, Muller I, Mullikin JC, Nguyen N, Nordsieck G, Nyakatura G, O'Dell CN, Okwuonu G, Palmer S, Pandian R, Parker D, Parrish J, Pasternak S, Patel D, Pearce AV, Pearson DM, Peulan SE, Perez L, Porter KM, Ramsey Y, Reichwald K, Rhodes S, Ridley EC, Schlessinger D, Schueler MG, Sekhah HK, Shaw-Smith C, Shen H, Sheridan EM, Shownkeen R, Skuce CD, Smith ML, Soberan EC, Steiginger HE, Stewart CA, Storey R, Swan RM, Swarbuck D, Tabor PE, Taudien S, Taylor T, Teague B, Thomas K, Thorpe A, Timms K, Tracey A, Trevorson S, Traversari AC, d'Urso R, Uyama H, Valenzano D, Waldron L, Wall M, Wang Q, Warren J, Wang GL, Wei X, West A, Whitehead SL, Whitley MN, Wilkinson JE, Willey DL, Williams G, Williams L, Williamson A, Williamson H, Wilming L, Woodmansey RL, Wray PW, Yen J, Zhang J, Zhou J, Zoghbi H, Zorilla S, Buck D, Reinhardt R, Poulikos A, Rosenbhal A, Lehrrach H, Meind A, Minot P, Hillier LW, Willard HF, Wilson RK, Waterston RH, Rice CM, Vaudin M, Coulson A, Nelson DL, Weinstock G, Sulton JE, Durbain R, Hubbard T, Gibbs RA, Beck S, Rogers J, Bentley DE: The DNA sequence of the human X chromosome. Nature 2005, 434(7031):325-337.

5. Pers H, Kong GH, Lee SW: MAGE-A1 overexpression reduces TNF-alpha cytotoxicity in ME-180 cells. Mol Cells 2002, 14(1):122-129.

6. Duan Z, Duan Y, Lamendola DE, Yusuf RZ, Naemen R, Persson RT, Seid MV: Overexpression of MAGE/GAGE genes in paclitaxel-resistant human cancer cell lines. Clin Cancer Res 2003, 9(7):2778-2785.

7. Glynn SA, Gammell P, Heenan M, O'Connor R, Liang Y, Keenan J, Clynnes M: A new suppressive in vitro phenotype induced by selection of human breast carcinoma cells with the chemo-immune drugs paclitaxel and doxorubicin. Br J Cancer 2004, 91(10):1800-1807.

8. Cilensek ZM, Yeheily F, Kular RK, Deiss LP: A member of the GAGE family of tumor antigens is an anti-apoptotic gene that confers resistance to Fas/CD95/APO-1, Interferon-gamma, taxol and gamma-irradiation. Cancer Biol Ther 2002, 1(4):380-387.

9. Yang B, O'Herrin S, Wu J, Reagan-Shaw M, Yu M, Nihal M, Longley BJ: Select Cancer Testis Antigens of the MAGE-A, -B, and -C Families Are Expressed in Mast Cell Lines and Promote Cell Viability In Vitro and In Vivo. J Invest Dermatol 2006.

10. Perry GH, Tchinda J, McGrath SD, Zhang J, Picker SR, Caceres AM, Iafraje A, Tyler-Smith C, Scherer SW, Eichler EE, Stone AC, Lee C: Hotspots for copy number variation in chimpanzees and humans. Proceedings of the National Academy of Sciences of the United States of America 2006, 103(2):8006-8011.

11. Bailey JA, Eichler EE: Primate segmental duplications: crucibles of evolution, diversity and disease. Nature reviews 2006, 7(7):552-564.

12. Patterson N, Richter DJ, Njerve S, Landes ES, Reich D: Genetic evidence for complex speciation of humans and chimpanzees. Nature 2006, 441(7079):1103-1108.

13. The Chimpanzee Sequencing and Analysis Consortium: Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 2005, 437(7055):69-87.

14. Emerson JJ, Kaessmann H, Betran E, Long M: Extensive gene traffic and amplification in African great apes and hominids. Genome research 2007, 17(13):1914-1918.

15. Chomez P, De Bakker O, Bertrand O, Deplaan E, Boon T, Lucas S: A testicular antigen gene family of cancer/testis-specific antigens: rapid evolution and amplification in African great apes and hominids. Proc Natl Acad Sci USA 2004, 101(22):7940-7945.
BMC Genomics 2007, 8:129

Proceedings of the National Academy of Sciences of the United States of America 2004, 101(9):3077-3082.

28. Birtle Z, Goodstadt L, Ponting C: Duplication and positive selection among hominin-specific PRAME genes. BMC genomics 2005, 6:120.

29. Kouprina N, Pavlicek A, Noskov VN, Solomon G, Otstotskij I, Isaacs W, Carpent JD, Trent JM, Schleuter J, Barrett JC, Jurka J, Lario NV: Dynamic structure of the SPANX gene cluster mapped to the prostate cancer susceptibility locus HPCX at Xq27. Genome research 2005, 15(11):1477-1486.

30. Gjerstorff MF, Kock K, Nielsen O, Ditlzel HJ: MAGE-A1, GAGE and NY-ESO-1 cancer/testis antigen expression during human gonadal development. Hum Reprod 2007.

31. Nelson PT, Zhang P, Spagnoli GC, Tomaszewski JE, Pasha TL, Frosina D, Caballero OL, Simpson AJ, Old LJ, Jungbluth AA: Cancer/testis (CT) antigens are expressed in fetal ovary. Cancer Immun 2007, 7:1.

32. Bustamante CD, Fledel-Alon A, Williamson S, Nielsen R, Hubisz MT, Glanowski S, Tanenbaum DM, White TJ, Sninsky JJ, Hernandez RD, Civello D, Adams MD, Cargill M, Clark AG: Natural selection on protein-coding genes in the human genome. Nature 2005, 437(7062):1153-1157.

33. Khahtovich P, Hellmann I, Enard W, Nowick K, Leinweber M, Franz H, Weiss G, Lachmann M, Paabo S: Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. Science 2005, 309(5742):1850-1854.

34. Khahtovich P, Enard W, Lachmann M, Paabo S: Evolution of primateogene expression. Nature reviews 2006, 7(9):693-702.

35. Nielsen R, Bustamante C, Hellmann I, Tanenbaum DM, White TJ, Sninsky JJ, Hernandez RD, Civello D, Adams MD, Cargill M, Clark AG: Natural selection on protein-coding genes in the human genome. Nature 2005, 437(7062):1153-1157.

36. Vicoso B, Charlesworth B: Evolution on the X chromosome: unusual patterns and processes. Nature reviews 2006, 7(8):645-653.

37. NCBI Reference Sequence (RefSeq) [http://www.ncbi.nlm.nih.gov/RefSeq]

38. CT Gene Database [http://www.cancerimmunity.org/CTdatabase]

39. Kasprzyk A, Keefe D, Smedley D, London D, Spooner W, Melsopp C, Hammond M, Rocca-Serra P, Cox T, Birney E: EnsMart: a generic system for fast and flexible access to biological data. Genome research 2004, 14(1):160-169.

40. BioMart - MartView [http://www.biomart.org/biomart/martview]

41. SIBsim4 project [http://sibsim4.sourceforge.net]

42. Thompson JD, Higgins DG, Gibson TJ: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic acids research 1994, 22(22):4673-4680.

43. R Development Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; 2004.

Publish with BioMed Central and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."
Sir Paul Nurse, Cancer Research UK

Your research papers will be:
- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

Page 11 of 11
(page number not for citation purposes)