

RESEARCH ARTICLE

Open Access

# Evidence of positive selection at codon sites localized in extracellular domains of mammalian CC motif chemokine receptor proteins

Kelsey J Metzger\*<sup>1,2</sup> and Michael A Thomas<sup>1</sup>

## Abstract

**Background:** CC chemokine receptor proteins (CCR1 through CCR10) are seven-transmembrane G-protein coupled receptors whose signaling pathways are known for their important roles coordinating immune system responses through targeted trafficking of white blood cells. In addition, some of these receptors have been identified as fusion proteins for viral pathogens: for example, HIV-1 strains utilize CCR5, CCR2 and CCR3 proteins to obtain cellular entry in humans. The extracellular domains of these receptor proteins are involved in ligand-binding specificity as well as pathogen recognition interactions.

In mammals, the majority of chemokine receptor genes are clustered together; in humans, seven of the ten genes are clustered in the 3p21-24 chromosome region. Gene conversion events, or exchange of DNA sequence between genes, have been reported in chemokine receptor paralogs in various mammalian lineages, especially between the cytogenetically closely located pairs CCR2/5 and CCR1/3. Datasets of mammalian orthologs for each gene were analyzed separately to minimize the potential confounding impact of analyzing highly similar sequences resulting from gene conversion events.

Molecular evolution approaches and the software package Phylogenetic Analyses by Maximum Likelihood (PAML) were utilized to investigate the signature of selection that has acted on the mammalian CC chemokine receptor (*CCR*) gene family. The results of neutral vs. adaptive evolution (positive selection) hypothesis testing using Site Models are reported. In general, positive selection is defined by a ratio of nonsynonymous/synonymous nucleotide changes ( $dN/dS$ , or  $\omega$ )  $> 1$ .

**Results:** Of the ten mammalian CC motif chemokine receptor sequence datasets analyzed, only *CCR2* and *CCR3* contain amino acid codon sites that exhibit evidence of positive selection using site based hypothesis testing in PAML. Nineteen of the twenty codon sites putatively identified as likely to be under positive selection code for amino acid residues located in extracellular domains of the receptor protein products.

**Conclusions:** These results suggest that amino acid residues present in intracellular and membrane-bound domains are more selectively constrained for functional signal transduction and homo- or heterodimerization, whereas amino acid residues in extracellular domains of these receptor proteins evolve more quickly, perhaps due to heightened selective pressure resulting from ligand-binding and pathogen interactions of extracellular domains.

## Background

Chemotactic or chemoattractant cytokine (chemokine) proteins are a unique division of cytokines characterized by their roles in cell signaling through the use of het-

erotrimeric GTP-binding (G protein-coupled) 7-transmembrane receptors [1-3]. Chemokines are the largest family of cytokines [4]. Currently, 42 ligand molecules and 19 receptors belong to the chemokine superfamily of cytokines [5]. Within the chemokine superfamily, protein sub-families are distinguished by differences in amino acid sequence motif of four conserved cysteines residues [6]. Chemokine ligands in the CC ( $\beta$ ) sub-family have two

\* Correspondence: kmetzger@umn.edu

<sup>1</sup> Department of Biological Sciences, Idaho State University, 921 South 8th Avenue, Campus Box 8007, Pocatello, ID 83209, USA  
Full list of author information is available at the end of the article

adjacent cysteines that are both involved in intra-chain disulfide bridges [7]. There are ten CC motif chemokine receptors (CCR1 through CCR10) that bind these ligands with differing ligand binding specificity [2,5].

Chemokines, in conjunction with adhesion molecules, recruit specific subpopulations of leukocytes by activating various receptors and initiating signaling through G protein-coupled pathways [8,9]. The signaling of chemokines through their receptors affect various cellular outcomes including leukocyte trafficking, gene transcription, degranulation of immune response cells, mitogenic processes, and apoptosis [2,5,6,8]. Chemokines generally act as secondary pro-inflammatory mediators that are induced by primary pro-inflammatory mediators [1].

Nucleotide mutations in the open reading frame coding for chemokine receptors can have a dramatic effect on receptor activity or little effect, depending on the location of the substitution and the nature of the amino acid replacement: amino acid substitutions resulting in alterations at key ligand binding extracellular domains or intracellular G-protein coupled domains in particular are known to result in disrupted or abnormal receptor activity [4]. Deficient signal transduction can result in increased susceptibility to infectious diseases as a result of the lack of a robust signaling response to pathogenic infection [10-16]. Mutations in the regulatory nucleotide sequence of chemokine receptors can also result in changes in gene expression and subsequent protein activity [17,18].

In addition to their roles as pro-inflammatory agents in innate immune response through binding of endogenous ligands and subsequent receptor activation, some chemokine receptors have been co-opted to act as receptors or fusion proteins for a number of pathogens including HIV-1 strains [19,20], protozoan parasites *Plasmodium knowlesi* and *P. vivax* [21], and Epstein-Barr virus [22]; herpesviruses mimic host chemokine receptors to elude host immune responses [23]. Mutations in regulatory or coding sequences for chemokine receptors that are used as pathogen fusion proteins can alter host-pathogen interactions: amino acid substitutions in extracellular domains can prevent recognition by the pathogen or interfere with the pathogen's ability to utilize the receptor as a gateway into the cell. Genetic markers associated with disease resistance have been found in regulatory and coding sequences of chemokine receptors [24-29].

Because of their critical role in signaling immune responses, chemokine receptors are subjects of intense selection to accommodate signaling molecules, and are expected to experience purifying selection to maintain conformation and functionality of ligand binding and signaling [30]. However, because of their role as targets of

pathogen entry, chemokine receptors are also expected to experience positive selection pressure in response to viral/pathogen hijacking [31]. Loci that are involved in responses to a variety of pathogens may experience balancing selection as a result of diverging selection pressures acting simultaneously on genes which may result in the maintenance of polymorphism [32]. Given this apparent evolutionary tension, it is of interest to investigate the signature of selection on this subfamily of chemokine receptors.

## Results

### Hypothesis Testing with Site Models

This paper presents the results of PAML hypothesis testing on chemokine receptor sequence datasets with "Site" models only [33,34]. These models analyze sequence data at the level of the codon, and test whether a hypothesis (model) that allows for positive selection ( $dN/dS > 1$  for some codons) is better fit to the data when compared to a null neutral hypothesis (model), determined through performing a likelihood ratio test between the likelihood scores of the null neutral and selection models. Each set of orthologous gene sequences was analyzed independently of one another such that only *CCR1* gene sequences were included in the first data set, only *CCR2* gene sequences were included in the second data set, and so on for each of the ten CCR genes (i.e., paralogous genes were not in the same data set).

When testing the hypothesis that some codon sites within chemokine receptor coding sequences have experienced positive selection pressure, significant results were obtained for some codons within the genes *CCR2* and *CCR3* (Table 1). For *CCR2*, the comparison between a null neutral site model which does not allow positive selection (M1a) and a selection site model (M2a) yielded a likelihood ratio test statistic of 5.45, which did not allow for rejection of the null hypothesis of neutral selection. However, the comparison between an additional pair of site models, M7 (null, neutral) and M8 (selection) for *CCR2* yielded a likelihood test ratio statistic of 15.22, significant at  $p = 0.001$ , and a proportion of sites (0.02430, 2.4%) with  $\omega = 2.83526$ . The analysis for *CCR2* had a total of 380 amino acid sites, and nine amino acid sites were identified as sites of positive selection (Table 2) using Bayes empirical Bayes (BEB) analysis [35]. One of the nine amino acid sites identified under positive selection had strong support with BEB posterior probability >95%. Eight of the nine amino acids identified as having experienced positive selection in the coding sequence of *CCR2* are located in extracellular domains of the protein; one positively selected amino acid residue is located in the second transmembrane domain, near the transmembrane/extracellular boundary (Figure 1).

**Table 1: Model Parameter Estimates, dN/dS Ratios, Log Likelihood Values and Test Statistics for PAML Site Models.**

Gene	Model	Parameters	dN/dS	p	l	2Δl
CCR2	M7: Neutral, beta	p = 0.33949, q = 1.05276	0.2417	2	-4931.71	
	M8: Selection, beta + ω	p <sub>0</sub> = 0.97570, p = 0.45144 q = 1.67059, (p <sub>1</sub> = 0.02430) ω = 2.83526	0.2728	4	-4924.10	M7 vs. M8: <b>15.22***</b>
CCR3	M1a: Nearly Neutral	ω <sub>0</sub> = 0.07487, ω <sub>1</sub> = 1.00 p <sub>0</sub> = 0.61460, (p <sub>1</sub> = 0.38540)	0.4314	2	-5536.21	
	M2a: Selection	ω <sub>0</sub> = 0.07422, ω <sub>1</sub> = 1.00 ω <sub>2</sub> = 5.44115, p <sub>0</sub> = 0.60315 p <sub>1</sub> = 0.38339, (p <sub>2</sub> = 0.01346)	0.5014	4	-5526.29	M1a vs. M2a: <b>19.84***</b>
CCR3	M7: Neutral, beta	p = 0.22415, q = 0.35810	0.3850	2	-5534.05	
	M8: Selection, beta + ω	p <sub>0</sub> = 0.98363, p = 0.23871 q = 0.3298, (p <sub>1</sub> = 0.01637) ω = 4.50735	0.2662	4	-5522.90	M7 vs. M8: <b>22.30***</b>

The test statistic 2Δl is compared to a  $\chi^2$  distribution with 2 degrees of freedom, critical values 5.99, 9.21, and 13.82 at 5%, 1%, and 0.1% significance, respectively. Significant results are indicated by asterisks. Values for non-significant comparisons are presented in Additional File 1.

For hypothesis testing of site models with CCR3 gene sequences, the comparison of M1a (neutral) vs. M2a (selection) site models led to rejection of the null, neutral hypothesis in favor of selection with a likelihood test ratio statistic of 19.84, significant at  $p = 0.001$ , with a proportion of sites (0.01346 1%), and  $\omega = 5.44115$ , indicating positive selection for those codon sites (Table 1). Five specific codon sites within the coding sequence were reported under positive selection with a BEB analysis following site testing: one site had strong support with BEB posterior probability greater than 95%, and one site had very strong support with BEB posterior probability greater than 99% (Table 2). A significant result was also obtained in the comparison of the second set of site models, M7 versus M8, with a test statistic of 22.3, significant at  $p = 0.001$ , with a proportion of sites (0.01637, ~1.6%) having a dN/dS value of 5.44 (Table 1). The analysis for CCR3 included 361 amino acid sites, and eleven specific amino acid sites were reported under positive selection with a BEB analysis following the site test. Three of these positively selected amino acid sites had posterior probabilities greater than 95%, one site had a posterior probability greater than 99% (Table 2). All of the positively

selected sites identified for CCR3 in this analysis were located in extracellular domains of the receptor protein (Figure 2). A summary of results (significant and non significant) for all site model tests is presented in Additional File 1.

## Discussion

None of the ten CC motif chemokine receptors has a signature of positive selection as indicated by an  $\omega$  value (ratio of nonsynonymous substitutions/synonymous substitutions, dN/dS) greater than one averaged over all codons as determined by hypothesis testing using branch models in PAML (unpublished results). Site tests, which analyze the sequence at the unit of the codon, revealed a proportion of codon sites that display evidence of positive selection ( $\omega > 1$ ) within the coding sequences of CCR2 and CCR3. The results obtained under the two sets of site models (M1a vs. M2a and M7 vs. M8) differ in some aspects; for example, the more conservative M1a vs. M2a comparison did not reveal statistically significant results for CCR2 while the M7 vs. M8 comparison did reveal significant differences, allowing for the identification of positively selected sites. For CCR3, while both M1a vs. M2a

**Table 2: Positively Selected Sites Under Different PAML Site Models Using Bayes Empirical Bayes Analysis.**

Gene	Model	Codon	Domain	Amino Acid	Posterior Probability	Post Mean +- SE for $\omega$
CCR2	M8: Selection, beta+ $\omega$	16	EC	S	0.632	1.536 +- 0.783
		23	EC	F	0.795	1.796 +- 0.781
		43	EC	Q	0.0916	1.977 +- 0.729
		95	MB	L	0.795	1.807 +- 0.810
		115	EC	L	0.623	1.535 +- 0.825
		183	EC	K	0.0926	1.986 +- 0.720
		187	EC	V	0.705	1.678 +- 0.831
		196	EC	R	0.961*	2.028 +- 0.691
		197	EC	G	0.862	1.910 +- 0.771
CCR3	M2a: Selection	4	EC	S	0.638	3.459 +- 2.147
		5	EC	L	0.988*	4.851 +- 1.496
		20	EC	V	0.998**	4.881 +- 1.453
		179	EC	T	0.919	4.581+- 1.731
		181	EC	T	0.878	4.378 +- 1.807
CCR3	M8: Selection, beta+ $\omega$	4	EC	S	0.824	2.296 +- 0.904
		5	EC	L	0.988*	2.590 +- 0.683
		20	EC	V	0.997**	2.603 +- 0.667
		31	EC	A	0.549	1.734 +- 0.921
		95	EC	R	0.681	2.000 +- 0.947
		96	EC	G	0.598	1.830 +- 0.940
		104	EC	G	0.863	2.347 +- 0.833
		177	EC	L	0.528	1.688 +- 0.902
		179	EC	E	0.550	1.748 +- 0.967
		181	EC	T	0.962*	2.545 +- 0.729

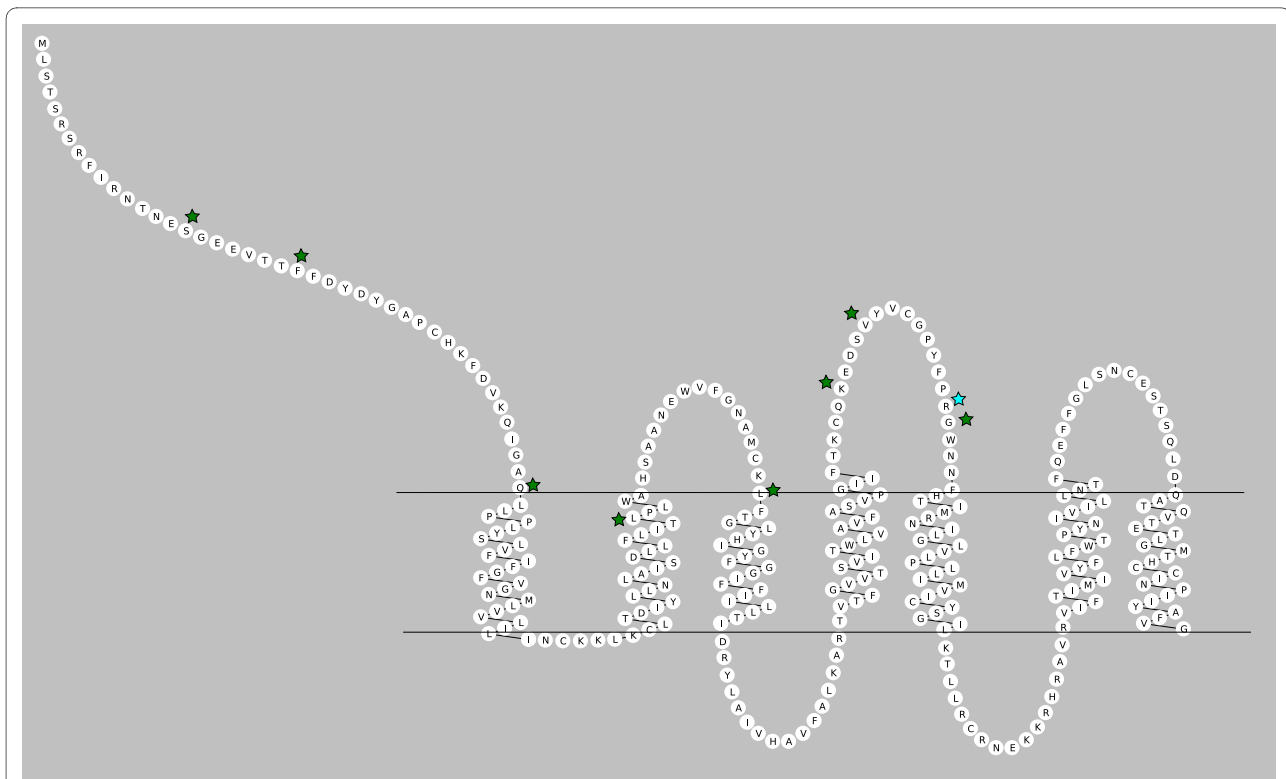
Asterisks indicate posterior probability  $P > 95\%$  (\*) and  $P > 99\%$ (\*\*). For codon position, the amino acid number is given followed by an abbreviation for the domain in which the amino acid is found: EC = extracellular, MB = membrane bound, IC = intracellular.

and M7 vs. M8 were both statistically significant comparisons, the comparison between M7 and M8 identified the same five codon sites that had been identified under M1a vs. M2a comparison as well as additional positively selected sites that were not identified in the M1a vs. M2a comparison. The differences in the results obtained using different models reflect that the M1a vs. M2a comparison is a more conservative test which may fail to detect positively selected sites identified by the less conservative M7 vs. M8 comparison.

It is interesting to note that in the results obtained for CCR2 and CCR3, nineteen out of the twenty amino acid sites that are identified as having experienced positive selection are located in the extracellular domains of the chemokine receptor proteins, suggesting that nonsynonymous substitutions are occurring, and more often being selected for, in the ligand binding and pathogen interac-

tion regions of the receptors. Previous studies on CCR2, CCR3 and other CC chemokine receptors have identified the amine-terminus and extracellular domains as being important for both the endogenous ligand-binding functions [36-42] as well as for binding efficacy for pathogens in situations where these receptors have been co-opted as fusion proteins [24,43-45]. These previous results coupled with the findings presented here point to the extracellular domains of CC chemokine receptor proteins being especially relevant to studies of the evolution of structure and function of receptors for endogenous ligand binding ability and as targets of pathogen interaction.

Recombination or gene conversion between paralogs in the chemokine receptor family has been investigated and described, particularly for CCR2/5 conversion in several orders of mammals [31,46-50]. Conversion events



**Figure 1 Location of CCR2 Positively Selected Codon Sites in Extracellular Receptor Protein Domains.** Stars indicate the location of positively selected sites in CCR2: Green stars indicate positively selected sites with a Bayes Empirical Bayes posterior probability  $\leq 95\%$ , blue stars indicate positively selected sites with posterior probability  $\geq 95\%$ . Diagram created with RbDe online software application [62].

between CCR1/3 in rodents have also been reported [31]. These conversion events have primarily involved transmembrane regions, but some conversion events have occurred more extensively over CCR gene sequences and have impacted extracellular domains; for example the first extracellular loop of CCR5 converted by recombination with CCR2 in *Mus* [31], extracellular loop 2 of CCR5 converted by recombination with CCR2 in *Homo* and *Oryctolagus* [31,46,50], and extracellular loop 3 of CCR3 converted by recombination with CCR1 in *Mus* [31].

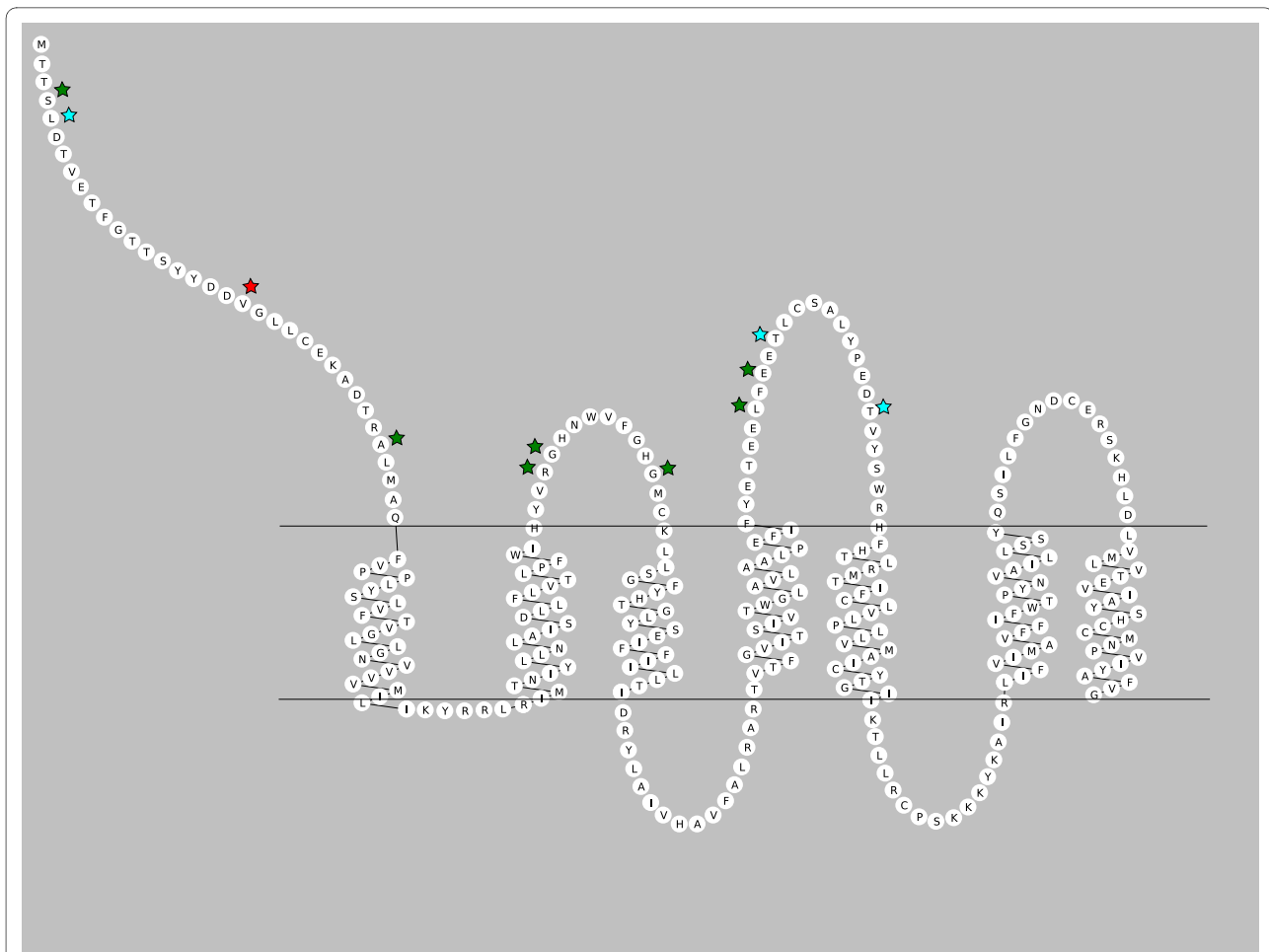
Analysis of sequences that have undergone gene conversion can lead to higher rate of false-positives when using maximum likelihood methods to detect positive selection, particularly in small data sets with only a few sequences, although the rate of false positives is only increased moderately [51]. To minimize the impact of gene conversion events on the results of this study and as an alternative to eliminating sequences or parts of sequences that have undergone conversion events, each set of orthologous genes was analyzed independently of one another such that highly similar sequences resulting from gene conversion events within a species involving paralogous genes were not included in the same analysis, but rather were analyzed independently in separate data sets (e.g., only CCR1 gene sequences were analyzed

together in one data set; CCR2 gene sequences were analyzed in a separate data set, and so on). In addition, a Bayes Empirical Bayes (BEB) analysis was used rather than Naïve Empirical Bayes (NEB) to identify putative codons under positive selection as NEB is less conservative and can be more prone to error in smaller data sets [34,35] whereas BEB produces a low rate of false-positives with sequences that have experienced gene conversion [51].

Both of the genes containing positively selected codon sites (CCR2 and CCR3) have been reported to have undergone gene conversion events as discussed above; however, in the case of CCR2, gene conversion events have led to the conversion of CCR5 by CCR2, whereas in the case of CCR3, it is CCR3 that has been converted by CCR1. The results presented here, in which only one gene of a gene conversion pair displays evidence of positive selection through hypothesis testing, indicate that independent analyses of sequences that have undergone gene conversion may mitigate the detection of false-positives due to gene conversion.

## Conclusions

Site tests, which analyze genetic sequences at the unit of the codon, revealed a proportion of codon sites that dis-



**Figure 2 Location of CCR3 Positively Selected Codon Sites in Extracellular Receptor Protein Domains.** Stars indicate the location of positively selected sites in CCR3: Green stars indicate positively sites with a Bayes Empirical Bayes posterior probability  $\leq 95\%$ , blue stars indicate positively sites with posterior probability  $\geq 95\%$ , red stars indicate sites under positive selection with posterior probability of  $\geq 99\%$ . Diagram created with RbDe online software application [62].

play evidence of positive selection ( $\omega > 1$ ) within the coding sequences of mammalian CC motif chemokine receptor genes *CCR2* and *CCR3*. Nineteen of the twenty amino acid sites identified as having experienced positive selection are located in extracellular domains of the chemokine receptor proteins *CCR2* and *CCR3*. These results suggest that amino acid residues present in intracellular and membrane-bound domains of mammalian CC motif chemokine receptor proteins are more selectively constrained, whereas amino acid residues in extracellular domains of these receptor proteins evolve more quickly, perhaps due to heightened selective pressure resulting from ligand-binding and pathogen interactions of extracellular domains.

## Methods

Genomic coding sequences for CCRs from a number of placental mammals were obtained through searches of

the online database NCBI Gene [52]. Taxa included in the data set were chosen using the recently updated placental mammal phylogeny [53] and the online software application TimeTree [54,55] to estimate divergence times between taxa (estimates given for nuclear genes were used). For PAML analyses, taxa that have diverged more than  $\sim 100$  MYA may lead to decreased analytical power due to highly divergent sequences, difficulty in sequence alignment, and saturation of substitutions [56-58]; therefore, only taxa that have diverged less than 100 MYA were included in the data set.

For genes that displayed alternative splicing patterns, the presumed ancestral isoform sequence was identified through alignment methods and included in the dataset while the alternative isoforms were not. For *CCR2*, the human isoform that localizes to the plasma membrane was included in the *CCR2* dataset, while the cytoplasmic variant was not. Species and GenBank accession numbers

for sequences used in the analyses are listed in Additional File 2.

Nucleotide alignments of chemokine receptor sequences were generated using amino acid sequence alignments and the software program TranAlign [59]. The output from TranAlign was converted to Nexus/PAUP format and submitted to the software program ModelTest [60] for selection of the most appropriate model of evolution for each dataset by testing the fit of 56 different evolutionary models with the data set. ModelTest uses both hierarchical likelihood ratio testing and Akaike Information Criterion (AIC). The best model was chosen based on AIC score and the number of estimated parameters. If there was a statistically insignificant difference between two models, the model with the fewest number of estimated parameters was chosen to introduce the least amount of uncertainty to the evolutionary analyses. Models used for analyses are summarized in Additional File 3.

#### Phylogenetic Analysis Using Parsimony\* (PAUP\*) and Phylogenetic Analysis by Maximum Likelihood (PAML) Methods

Maximum likelihood phylogenetic trees for each data set (Additional Files 4, 5, 6, 7, 8, 9, 10, 11, 12 and 13) were constructed with the software package PAUP\* [61]. Data sets and maximum likelihood phylogenetic trees for each gene were submitted to PAML CODEML version 4.1 under different models and parameters to test for adaptive evolution either at codon sites ("Site Model"), along lineages ("Branch Model"), or at sites within lineages ("Branch-Site Model") [34]. This paper presents the results of testing the data sets described with "Site" models only.

"Site" models allow the dN/dS ratio to vary across codons within a sequence for a lineage. Proportions of sites within each lineage were estimated to be in different categories: positive selection is indicated by some codons having a dN/dS > 1. The null, neutral model does not allow positive selection and is compared to the alternative hypothesis in which in which positive selection is allowed.

Two sets of site models are commonly used to test hypotheses of selection, and have been used here: M1a vs. M2a and M7 vs. M8. In the first set of models, the model M1a: Nearly Neutral allows 2 categories of codon sites in  $p_0$ , and  $p_1$  proportions, with  $\omega_0 < 1$ , and  $\omega_1 = 1$ , whereas the model M2a: Selection allows an additional category of codons ( $p_2$ ) with  $\omega_2 > 1$ , indicating positive selection. The second set of site models compared is M7 and M8, in which M7 specifies a neutral model with dN/dS ratios across a continuous beta distribution with estimated parameters  $p$  and  $q$  of the beta distribution, and M8 spec-

ifies a similar model with an additional category for sites that have dN/dS > 1, indicating positive selection. M7 assumes a beta distribution of  $\omega$  values between 0 and 1, and therefore does not allow any sites under positive selection ( $\omega > 1$ ). The M8 model is similar to M7 in that it also assumes a beta distribution for omega values, but allows another category of sites in which  $\omega > 1$ . The comparison between M7: beta and M8: beta +  $\omega$  is less conservative, and may indicate positive selection even when none is detected by the M1a: M2a comparison.

The PAML settings for the null (neutral) model M1a were model = 0, NSsites = 1, and for the alternative (selection) model M2a were model = 0, NSsites = 2. The PAML settings for the null model M7 were model = 0, NSsites = 7, and for the alternative (selection) model M8 were model = 0, NSsites = 8.

The likelihood estimates for each were compared using a hierarchical Likelihood Ratio Test (hLRT) of twice the difference in log likelihood values of the models being compared ( $2\Delta\ln L$ ), with the result approximating chi-square distribution with degrees of freedom for the test statistic determined by the difference in estimated parameters between the models being compared. For both the M1a (neutral) vs. M2a (selection) and M7 (beta) vs. M8 (beta + selection) comparisons, the null model has two estimated parameters, while the alternative estimates four, resulting in two degrees of freedom and chi-square critical values of 5.99, 9.21, and 13.82 at 5%, 1%, and 0.1% significance, respectively [34].

#### Additional material

**Additional file 1 Supplementary Table 1: Model Parameter Estimates, dN/dS Ratios, Log Likelihood Values and Test Statistics for PAML Site Models.** Summary of results for all PAML hypothesis testing using site models. Includes significant and non-significant results.

**Additional file 2 Supplementary Table 2: Taxa and NCBI GenBank accession numbers for loci included in CC chemokine receptor data sets.** List of species and NCBI GenBank accession numbers for sequences used to construct the ten datasets (for each of the ten sets of orthologous genes) for hypothesis testing. Species and accession numbers for each dataset are grouped together on the table.

**Additional file 3 Supplementary Table 3: Parameters for ModelTest models of evolution used in PAML hypothesis testing of CC chemokine receptor sequences.** Summary of evolutionary model parameters used in PAML hypothesis testing.

**Additional file 4 Supplementary Figure 1: Maximum likelihood phylogeny of mammalian CCR1 gene sequences.** A phylogeny of mammalian CCR1 gene sequences. The tree was produced using PHYML with the GTR nucleotide model, a discrete gamma model with four categories, and a shape parameter of 0.6525. Bootstrapping was performed with 100 replicates. Bootstrap support is indicated at nodes.

**Additional file 5 Supplementary Figure 2: Maximum likelihood phylogeny of mammalian CCR2 gene sequences.** A phylogeny of mammalian CCR2 gene sequences. The tree was produced using PHYML with the GTR nucleotide model, a discrete gamma model with four categories, and a shape parameter of 0.5867. Bootstrapping was performed with 100 replicates. Bootstrap support is indicated at nodes.

**Additional file 6 Supplementary Figure 3: Maximum likelihood phylogeny of mammalian CCR3 gene sequences.** A phylogeny of mammalian CCR3 gene sequences. The tree was produced using PHYML with the GTR nucleotide model, a discrete gamma model with four categories, proportion of invariable sites 0.3247 and a shape parameter of 2.8971. Bootstrapping was performed with 100 replicates. Bootstrap support is indicated at nodes.

**Additional file 7 Supplementary Figure 4: Maximum likelihood phylogeny of mammalian CCR4 gene sequences.** A phylogeny of mammalian CCR4 gene sequences. The tree was produced using PHYML with the GTR nucleotide model, a discrete gamma model with four categories, proportion of invariable sites 0.5901 and an estimated shape parameter of 33.284. Bootstrapping was performed with 100 replicates. Bootstrap support is indicated at nodes.

**Additional file 8 Supplementary Figure 5: Maximum likelihood phylogeny of mammalian CCR5 gene sequences.** A phylogeny of mammalian CCR5 gene sequences. The tree was produced using PHYML with the HKY nucleotide model, a discrete gamma model with four categories, and a shape parameter of 0.451. Bootstrapping was performed with 100 replicates. Bootstrap support is indicated at nodes.

**Additional file 9 Supplementary Figure 6: Maximum likelihood phylogeny of mammalian CCR6 gene sequences.** A phylogeny of mammalian CCR6 gene sequences. The tree was produced using PHYML with the JC69 nucleotide model, a discrete gamma model with four categories, proportion of invariable sites 0.4667, and an estimated shape parameter of 100. Bootstrapping was performed with 100 replicates. Bootstrap support is indicated at nodes.

**Additional file 10 Supplementary Figure 7: Maximum likelihood phylogeny of mammalian CCR7 gene sequences.** A phylogeny of mammalian CCR7 gene sequences. The tree was produced using PHYML with the GTR nucleotide model, a discrete gamma model with four categories, and a shape parameter of 0.257. Bootstrapping was performed with 100 replicates. Bootstrap support is indicated at nodes.

**Additional file 11 Supplementary Figure 8: Maximum likelihood phylogeny of mammalian CCR8 gene sequences.** A phylogeny of mammalian CCR8 gene sequences. The tree was produced using PHYML with the HKY nucleotide model, a discrete gamma model with four categories, and a shape parameter of 0.6482. Bootstrapping was performed with 100 replicates. Bootstrap support is indicated at nodes.

**Additional file 12 Supplementary Figure 9: Maximum likelihood phylogeny of mammalian CCR9 gene sequences.** A phylogeny of mammalian CCR9 gene sequences. The tree was produced using PHYML with the HKY nucleotide model, a discrete gamma model with four categories, and a shape parameter of 0.3547. Bootstrapping was performed with 100 replicates. Bootstrap support is indicated at nodes.

**Additional file 13 Supplementary Figure 10: Maximum likelihood phylogeny of mammalian CCR10 gene sequences.** A phylogeny of mammalian CCR10 gene sequences. The tree was produced using PHYML with the GTR nucleotide model, a discrete gamma model with four categories, and a shape parameter of 0.2439. Bootstrapping was performed with 100 replicates. Bootstrap support is indicated at nodes.

#### Authors' contributions

KM carried out dataset construction, molecular evolution analyses, and drafting of the manuscript. MT contributed to the conception and design of the study, and participated in critical review and revision of the manuscript. All authors read and approved the final manuscript.

#### Author Details

<sup>1</sup>Department of Biological Sciences, Idaho State University, 921 South 8th Avenue, Campus Box 8007, Pocatello, ID 83209, USA and <sup>2</sup>Center for Learning Innovation, University of Minnesota Rochester, 300 University Square, 111 South Broadway, Rochester, MN 55904, USA

Received: 21 October 2009 Accepted: 10 May 2010

Published: 10 May 2010

#### References

1. Graves DT, Jiang Y: Chemokines, a family of chemotactic cytokines. *Crit Rev Oral Biol Med* 1995, **6**:109-118.
2. Rosenkilde MM, Schwartz TW: The chemokine system - a major regulator of angiogenesis in health and disease. *APMIS* 2004, **112**:481-495.
3. Goncharova LB, Tarakanov AO: Why chemokines are cytokines while their receptors are not cytokine ones? *Curr Med Chem* 2008, **15**:1297-1304.
4. Fernandez EJ, Lolis E: Structure, function, and inhibition of chemokines. *Annu Rev Pharmacol Toxicol* 2002, **42**:469-499.
5. Bonecchi R, Galliera E, Borroni EM, Corsi MM, Locati M, Mantovani A: Chemokines and chemokine receptors: an overview. *Front Biosci* 2009, **14**:540-551.
6. Baggiolini M, Dewald B, Moser B: Human chemokines: An update. *Annu Rev Immunol* 1997, **15**:675-705.
7. Lodi PJ, Garrett DS, Kuszewski J, Tsang ML, Weatherbee JA, Leonard WJ, Gronenborn AM, Clore GM: High-resolution solution structure of the beta-chemokine HMIP-1-beta by multidimensional NMR. *Science* 1994, **263**:1762-1767.
8. Rossi D, Zlotnik A: The biology of chemokines and their receptors. *Annu Rev Immunol* 2000, **18**:217-242.
9. de Paz JL, Moseman EA, Noti C, Polito L, von Andrian UH, Seeberger PH: Profiling heparin-chemokine interactions using synthetic tools. *ACS Chem Biol* 2007, **2**:735-744.
10. Carrington M, Dean M, Martin MP, O'Brien SJ: Genetics of HIV-1 infection: chemokine receptor CCR5 polymorphism and its consequences. *Hum Mol Genet* 1999, **8**:1939-1945.
11. Lim JK, Glass WG, McDermott DH, Murphy PM: CCR5: No longer a 'good for nothing' gene - chemokine control of West Nile Virus infection. *Trends Immunol* 2006, **27**:308-312.
12. Glass WG, Lim JK, McDermott DH, Pletnev A, Cholera R, Gao J, Lekhong S, Yu SF, Frank WA, Pape J, Cheshier RC, Murphy PM: CCR5 saves lives: the protective role of CCR5 during West Nile virus infection. *J Neurochem* 2006, **96**:94-94.
13. Glass WG, McDermott DH, Lim JK, Lekhong S, Yu SF, Frank WA, Pape J, Cheshier RC, Murphy PM: CCR5 deficiency increases risk of symptomatic West Nile virus infection. *J Exp Med* 2006, **203**:35-40.
14. Navratilova Z: Polymorphisms in CCL2 and CCL5 chemokines/chemokine receptor genes and their association with diseases. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub* 2006, **150**:191-204.
15. Balestrieri ML, Balestrieri A, Mancini FP, Napoli C: Understanding the immunoangiostatic CXC chemokine network. *Cardiovasc Res* 2008, **78**:250-256.
16. Lim JK, Louie CY, Glaser C, Jean C, Johnson B, Johnson H, McDermott DH, Murphy PM: Genetic deficiency of chemokine receptor CCR5 is a strong risk factor for symptomatic West Nile Virus infection: A meta-analysis of 4 cohorts in the US epidemic. *J Infect Dis* 2008, **197**:262-265.
17. Mummid S, Bamshad M, Ahuja SS, Gonzalez E, Feuillet PM, Begum K, Galvis MC, Kostecki V, Valente AJ, Murthy KK, Haro L, Dolan MJ, Allan JS, Ahuja SK: Evolution of human and non-human primate CC chemokine receptor 5 gene and mRNA - Potential roles for haplotype and mRNA diversity, differential haplotype-specific transcriptional activity, and altered transcription factor binding to polymorphic nucleotides in the pathogenesis of HIV-1 and Simian immunodeficiency virus. *J Biol Chem* 2000, **275**:18946-18961.
18. Nadif R, Mintz M, Rivas-Fuentes S, Jedlicka A, Lavergne E, Rodero M, Kauffmann F, Combadiere C, Kleeberger SR: Polymorphisms in chemokine and chemokine receptor genes and the development of coal workers' pneumoconiosis. *Cytokine* 2006, **33**:171-178.
19. Alkhatib G, Combadiere C, Broder CC, Feng Y, Kennedy PE, Murphy PM, Berger EA: CC CKRS: A RANTES, MIP-1 alpha, MIP-1 beta receptor as a fusion cofactor for macrophage-tropic HIV-1. *Science* 1996, **272**:1955-1958.
20. Broder CC, Collman RG: Chemokine receptors and HIV. *J Leukocyte Biol* 1997, **62**:20-29.
21. Horuk R, Chitnis CE, Darbonne WC, Colby TJ, Rybicki A, Hadley TJ, Miller LH: A receptor for the malarial parasite *Plasmodium vivax* - the erythrocyte chemokine receptor. *Science* 1993, **261**:1182-1184.



22. Vischer HF, Nijmeijer S, Smit MJ, Leurs R: **Viral hijacking of human receptors through heterodimerization.** *Biochem and Biophys Res Commun* 2008, **377**:93-97.
23. Murphy PM: **Molecular piracy of chemokine receptors by herpesviruses.** *Infect Agents Dis Rev Issues Comment* 1994, **3**:137-154.
24. Howard OMZ, Shirakawa AK, Turpin JA, Maynard A, Tobin GJ, Carrington M, Oppenheim JJ, Dean M: **Naturally occurring CCR5 extracellular and transmembrane domain variants affect HIV-1 co-receptor and ligand binding function.** *J Biol Chem* 1999, **274**:16228-16234.
25. Singh KK, Barroga CF, Hughes MD, Chen J, Raskino C, McKinney RE, Spector SA: **Prevalence of chemokine and chemokine receptor polymorphisms in seroprevalent children with symptomatic HIV-1 infection in the United States.** *J Acquir Immune Defic Syndr* 2004, **35**:309-313.
26. Wang CB, Song W, Lobashevsky E, Wilson CM, Douglas SD, Mytilineos J, Schoenbaum EE, Tang JM, Kaslow RA: **Cytokine and chemokine gene polymorphisms among ethnically diverse North Americans with HIV-1 infection.** *J Acquir Immune Defic Syndr* 2004, **35**:446-454.
27. Kaur G, Singh P, Kumar N, Raptap CC, Sharma G, Vajpayee M, Wig N, Sharma SK, Mehra NK: **Distribution of CCR2 polymorphism in HIV-1 infected and healthy subjects in North India.** *Int J Immunogenet* 2007, **34**:153-156.
28. Suresh P, Wanchu A, Sachdeva RK, Bhatnagar A: **Gene polymorphisms in CCR5, CCR2, CX3CR1, SDF-1 and RANTES in exposed but uninfected partners of HIV-1 infected individuals in North India.** *J Clin Immunol* 2007, **27**:131-131.
29. Zhao WY, Lee SS, Wong KH, Chan KCW, Ng T, Chan CCS, Han D, Yam WC, Yuen KY, Ng MH, Zheng BJ: **Functional analysis of naturally occurring mutations in the open reading frame of CCR5 in HIV-infected Chinese patients and healthy controls.** *J Acquir Immune Defic Syndr* 2005, **38**:509-517.
30. Kunstman KJ, Puffer B, Korber BT, Kuiken C, Smith UR, Kunstman J, Stanton J, Agy M, Shibata R, Yoder AD, Pillai S, Doms RW, Marx P, Wolinsky SM: **Structure and function of CC-chemokine receptor 5 homologues derived from representative primate species and subspecies of the taxonomic suborders Prosimii and Anthroipoidea.** *J Virol* 2003, **77**:12310-12318.
31. Shields DC: **Gene conversion among chemokine receptors.** *Gene* 2000, **246**:239-245.
32. Hedrick PW: **Balancing selection.** *Curr Biol* 2007, **17**:R230-R231.
33. Yang Z: **PAML: a program package for phylogenetic analysis by maximum likelihood.** *Comput Appl BioSci* 1997, **13**:555-556.
34. Yang Z: **PAML 4: Phylogenetic Analysis by Maximum Likelihood.** *Mol Biol Evol* 2007, **24**:1586-1591.
35. Yang ZH, Wong WSW, Nielsen R: **Bayes empirical Bayes inference of amino acid sites under positive selection.** *Mol Biol Evol* 2005, **22**:1107-1118.
36. Pease JE, Wang J, Ponath PD, Murphy PM: **The N-terminal extracellular segments of the chemokine receptors CCR1 and CCR3 are determinants for MIP-1 alpha and eotaxin binding, respectively, but a second domain is essential for efficient receptor activation.** *J Biol Chem* 1998, **273**:19972-19976.
37. Han KH, Green SR, Tangirala RK, Tanaka S, Quehenberger O: **Role of the first extracellular loop in the functional activation of CCR2 - The first extracellular loop contains distinct domains necessary for both agonist binding and transmembrane signaling.** *J Biol Chem* 1999, **274**:32055-32062.
38. Zoffmann S, Chollet A, Galzi JL: **Identification of the extracellular loop 2 as the point of interaction between the n terminus of the chemokine MIP-1 alpha and its CCR1 receptor.** *Mol Pharmacol* 2002, **62**:729-736.
39. Blanpain C, Doranz BJ, Bondue A, Govaerts C, De Leener A, Vassart G, Doms RW, Proudfoot A, Parmentier M: **The core domain of chemokines binds CCR5 extracellular domains while their amino terminus interacts with the transmembrane helix bundle.** *J Biol Chem* 2003, **278**:5179-5187.
40. Sabroe I, Jorritsma A, Stubbs VEL, Xanthou G, Jopling LA, Panath PD, Williams TJ, Murphy PM, Pease JE: **The carboxyl terminus of the chemokine receptor CCR3 contains distinct domains which regulate chemotactic signaling and receptor down-regulation in a ligand-dependent manner.** *Eur J Immunol* 2005, **35**:1301-1310.
41. Duchesnes UE, Murphy PM, Williams TJ, Pease JE: **Alanine scanning mutagenesis of the chemokine receptor CCR3 reveals distinct extracellular residues involved in recognition of the eotaxin family of chemokines.** *Mol Immunol* 2006, **43**:1221-1231.
42. Rajagopalan L, Rajarathnam K: **Structural basis of chemokine receptor function - A model for binding affinity and ligand selectivity.** *Biosci Rep* 2006, **26**:325-339.
43. Frade JMR, Llorente M, Mellado M, Alcami J, GutierrezRamos JC, Zaballós A, delReal G, Martínez AC: **The amino-terminal domain of the CCR2 chemokine receptor acts as coreceptor for HIV-1 infection.** *J Clin Invest* 1997, **100**:497-502.
44. Liu SQ, Fan SX, Sun ZR: **Structural and functional characterization of the human CCR5 receptor in complex with HIV gp120 envelope glycoprotein and CD4 receptor by molecular modeling studies.** *J Mol Modeling* 2003, **9**:329-336.
45. Ho PT, Teal BE, Ross TM: **Multiple residues in the extracellular domains of CCR3 are critical for coreceptor activity.** *Virology* 2004, **329**:109-118.
46. Carmo CR, Esteves PJ, Ferrand N, Loo W van der: **Genetic variation at chemokine receptor CCR5 in leporids: alternation at the 2<sup>nd</sup> extracellular domain by gene conversion with CCR2 in *Oryctolagus*, but not in *Sylvilagus* and *Lepus* species.** *Immunogenetics* 2006, **58**:494-501.
47. Esteves PJ, Abrantes J, Loo W van der: **Extensive gene conversion between CCR2 and CCR5 in domestic cat (*Felis catus*).** *Int J Immunogenet* 2007, **34**:321-324.
48. Vazquez-Salat N, Yuhki N, Beck T, O'Brien SJ, Murphy WJ: **Gene conversion between mammalian CCR2 and CCR5 chemokine receptor genes: A potential mechanism for receptor dimerization.** *Genomics* 2007, **90**:213-224.
49. Pereygin AA, Zharkikh AA, Astakhova NM, Lear TL, Brinton MA: **Converted evolution of vertebrate CCR2 and CCR5 genes and the origin of a recombinant equine CCR5/2 gene.** *J Hered* 2008, **99**:500-511.
50. Abrantes J, Carmo CR, Matthee CA, Yamada F, Loo W van der, Esteves PJ: **A shared unusual genetic change at the chemokine receptor type 5 between *Oryctolagus*, *Bunolagus* and *Pentalagus*.** *Conserv Genet* 2009. DOI 10.1007/s10592-009-9990-1.
51. Casola C, Hahn MW: **Gene Conversion Among Paralogs Results in Moderate False Detection of Positive Selection Using Likelihood Methods.** *J Mol Evol* 2009, **68**:679-687.
52. **National Center for Biotechnology Information Entrez Gene** [<http://www.ncbi.nlm.nih.gov/gene>]
53. Springer MS, Burk-Herrick A, Meredith R, Eizirik E, Teeling E, O'Brien SJ, Murphy WJ: **The adequacy of morphology for reconstructing the early history of placental mammals.** *Syst Biol* 2007, **56**:673-684.
54. **TimeTree: The Timescale of Life** [<http://www.timetree.org>]
55. Hedges SB, Dudley J, Kumar S: **TimeTree: a public knowledge-base of divergence times among organisms.** *Bioinformatics* 2006, **22**:2971-2972.
56. Yang ZH: **On the best evolutionary rate for phylogenetic analysis.** *Syst Biol* 1998, **47**:125-133.
57. Anisimova M, Bielawski JP, Yang ZH: **Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution.** *Mol Biol Evol* 2001, **18**:1585-1592.
58. Bielawski JP, Yang ZH: **Maximum likelihood methods for detecting adaptive evolution after gene duplication.** *J Struct Funct Genomics* 2003, **3**:201-212.
59. Rice P, Longden I, Bleasby A: **EMBOSS: The European Molecular Biology Open Software Suite.** *Trends Genet* 2000, **16**:276-277.
60. Posada D, Crandall KA: **Modeltest: testing the model of DNA substitution.** *Bioinformatics* 1998, **14**:817-818.
61. Swofford DL: **PAUP\*. Phylogenetic Analysis Using Parsimony (\*and Other Methods).** Version 4 Sinauer Associates, Sunderland, MA; 2002.
62. Skrabanek L, Campagne F, Weinstein H: **Building protein diagrams on the web with the residue-based diagram editor RdBe.** *Nucleic Acids Res* 2003, **31**:3856-3858.

doi: 10.1186/1471-2148-10-139

**Cite this article as:** Metzger and Thomas, Evidence of positive selection at codon sites localized in extracellular domains of mammalian CC motif chemokine receptor proteins *BMC Evolutionary Biology* 2010, **10**:139