

# Adaptive Evolution of *cry* Genes in *Bacillus thuringiensis*: Implications for Their Specificity Determination

Jin-Yu Wu<sup>1</sup>, Fang-Qing Zhao<sup>2</sup>, Jie Bai<sup>1</sup>, Gang Deng<sup>1</sup>, Song Qin<sup>2</sup>, and Qi-Yu Bao<sup>1\*</sup>

<sup>1</sup>*Institute of Biomedical Informatics, Wenzhou Medical College, Wenzhou 325000, China;* <sup>2</sup>*Insititute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China.*

The *cry* gene family, produced during the late exponential phase of growth in *Bacillus thuringiensis*, is a large, still-growing family of homologous genes, in which each gene encodes a protein with strong specific activity against only one or a few insect species. Extensive studies are mostly focusing on the structural and functional relationships of Cry proteins, and have revealed several residues or domains that are important for the target recognition and receptor attachment. In this study, we have employed a maximum likelihood method to detect evidence of adaptive evolution in Cry proteins, and have identified 24 positively selected residues, which are all located in Domain II or III. Combined with known data from mutagenesis studies, the majority of these residues, at the molecular level, contribute much to the insect specificity determination. We postulate that the potential pressures driving the diversification of Cry proteins may be in an attempt to adapt for the “arm race” between  $\delta$ -endotoxins and the targeted insects, or to enlarge their target spectra, hence result in the functional divergence. The sites identified to be under positive selection would provide targets for further structural and functional analyses on Cry proteins.

**Key words:** adaptive evolution, *Bacillus thuringiensis*, Cry protein, maximum likelihood method

## Introduction

*Bacillus thuringiensis*, a naturally occurring Gram-positive bacterium, can produce parasporal crystal inclusions consisting of one or several  $\delta$ -endotoxins (Cry proteins), which have insecticidal properties affecting a selective range of insect orders (1). The *cry* gene family is a large, still-growing family of homologous genes, in which each gene encodes a protein with strong specific activity against only one or a few insect species. When the crystal proteins are ingested by insects, they are solubilized in the alkaline environment of the insect gut, releasing their constituent Cry proteins as protoxins. Midgut subsequently converts the protoxins into biologically active toxins by proteolytic enzymes. These activated toxins then bind to specific receptors on the surface of the midgut epithelial cells and insert into the cell membranes, forming pores and channels in the gut cell membrane, followed by destruction of the epithelial cells (2, 3). Receptor binding is the major determinant of host specificity by different Cry proteins. In view of such properties,

*B. thuringiensis* is developed as a type of microbial insecticide and has already been a useful alternative or supplement to synthetic chemical pesticide application on commercial agriculture, forest management, and mosquito control. It is also a key source of genes for pest-resistant transgenic plants (1).

To date, more than 150 different Cry toxins have been cloned and tested, and a new nomenclature has been formulated to accommodate the growing list of new toxin genes/proteins according to their amino acid sequence identities (4). Five tertiary structures of Cry proteins have been determined through X-ray crystallography, namely Cry1Aa, Cry2Aa, Cry3Aa, Cry3Bb, and Cry4Ba. They all suggest that the active toxins are globular molecules with three conserved domains. Domain I, comprising seven  $\alpha$ -helices, is thought to be responsible for inserting into insect cell membrane and to be involved in pore formation (5, 6). Domain II consists of three anti-parallel  $\beta$ -sheets, each terminating in surface-exposed loops, which are the most variable part in the Cry toxin structure. This domain has been demonstrated to participate in receptor recognition and hence deter-

**\*Corresponding author.**

**E-mail:** baoqy@genomics.org.cn

mines the insect specificity (1, 7). Domain III is made of two anti-parallel  $\beta$ -sheets into  $\beta$ -sandwich structure. The role of Domain III at the molecular level is still unknown, although a variety of mutagenesis experiments have shown that it can also be involved in receptor binding and specificity determination (8, 9).

In comparison to the extensive studies describing the function and the specificity determination regions, comprehensive evolutionary analyses of the *cry* gene family are rather rare. Positive selection is a phenomenon favoring the retention of mutations that are beneficial to an individual or population, and is thought to be an ephemeral case frequently resulting in the occurrence of a protein with novel function (10). The nonsynonymous to synonymous substitution rate ratio ( $\omega = d_N/d_S$ ) provides a sensitive and effective measure of selective pressure at the protein level, with  $\omega$  values  $< 1$ ,  $= 1$ , and  $> 1$  indicating purifying (negative) selection, neutral evolution, and positive selection, respectively. The specific goal of this study is to identify whether the *cry* gene family has been subjected to positive selection in the process of evolution by using the maximum likelihood models of codon substitution implemented in the PAML package (11). The major advantage of these models is that they can account for variable selective pressures by assuming a statistical distribution of the  $\omega$  ratio among sites (12). Here, the likelihood ratio test (LRT) is utilized for some clades of Cry proteins to study the evolution of the *cry* gene family and elucidate potential factors that drive the diversification of Cry proteins in *B. thuringiensis*. We believe that identification of the selection pressures acting on *cry* genes together with functional and structural data will shed light on their evolution processes and functional implications.

## Results

### Phylogenetic analysis

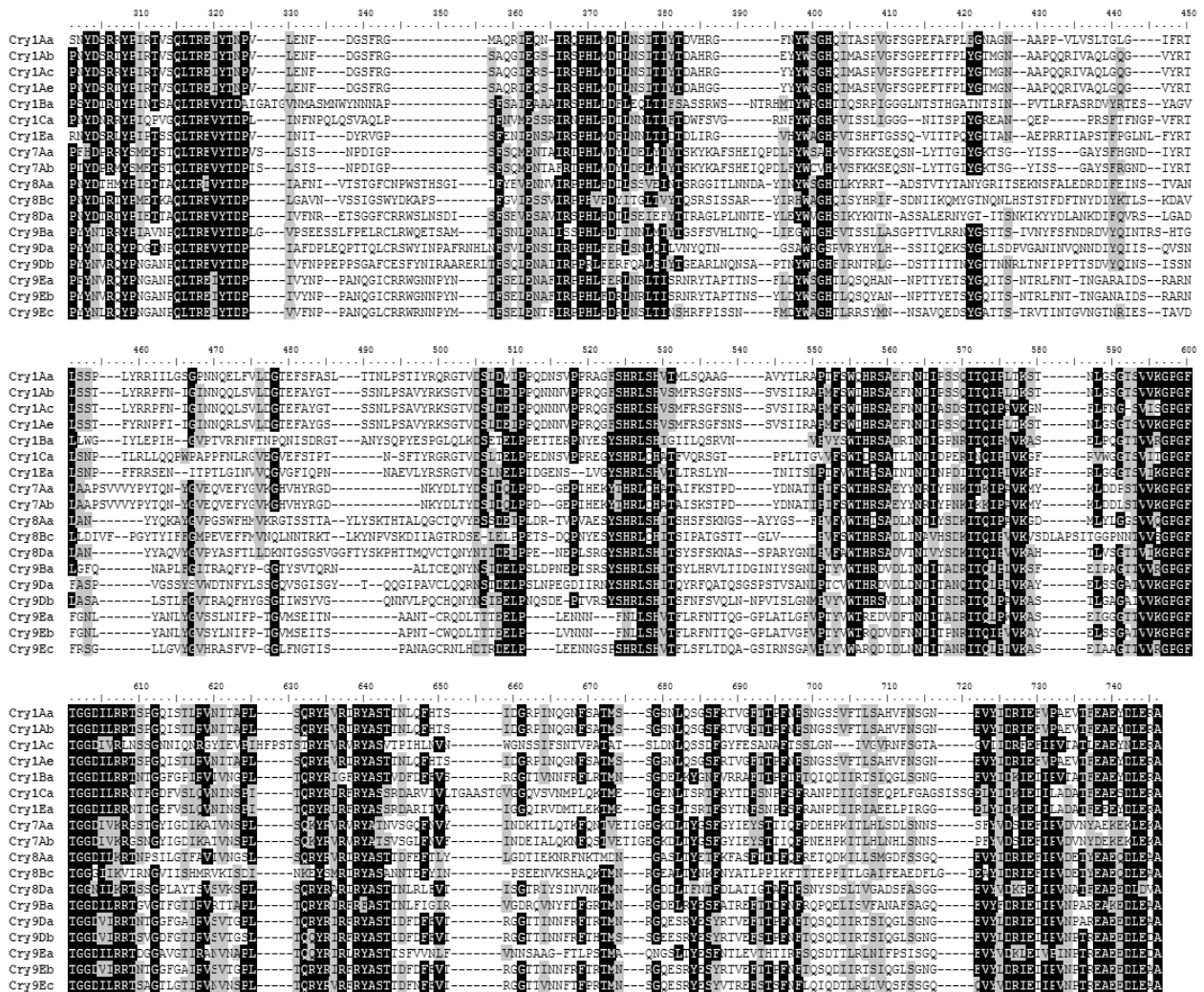
Before carrying out any tests of evolutionary models, we have examined the phylogenetic relationships of different subfamilies of *cry* genes. The sequence alignment result of 18 Cry proteins in Domains II and III is shown in Figure 1. Phylogenies constructed by neighbor-joining (NJ), minimum evolution (ME), and maximum parsimony (MP) methods exhibited quite similar topologies (data not shown), with high bootstrap value supports ( $> 90\%$ ) for each subfamily. Here we just show the phylogenetic tree constructed by the ME method (Figure 2). From this tree, we can

see that every subfamily is clustered into one group with a very high degree of sequence similarity and a bootstrap value of 100, indicating that these sequences might have evolved from a common ancestor.

### Maximum likelihood analysis of positive selection

The likelihood model analysis of the selection pressures acting on the Cry proteins (see Materials and Methods), which assumes variable selective pressures among sites, provided strong evidence for positive selection (Table 1). The one-ratio model (M0) gives a log likelihood score of  $-42,707.52$ , with the estimate  $\omega = 0.247$ . The log likelihood score under the “nearly neutral” model (M1a) is  $-41,394.82$ , with parameter estimates  $\omega_0 = 0.128$  and  $\omega_1 = 1.000$ . The “positive selection” model (M2a) suggests that about 3.7% of the sites are under positive selection with  $\omega_2 = 2.798$ . Because M2a is an extension of M1a (neutral), these two models can be compared by using LRT analysis. The test statistic is  $2\Delta \ln L = 49.80$ , with  $p < 0.001$  and degree of freedom (df) = 2. So M2a provides significantly better fit to the data than M1a. M3 also fits the data significantly better than M0 with  $p < 0.001$  and df = 3. The “ $\beta$  and  $\omega$ ” model (M8) suggests that about 12.5% of the sites are under positive selection with  $\omega = 1.468$ . We found that M8 fits the data significantly better than M7 ( $2\Delta \ln L = 182.30$ , df = 2,  $p < 0.001$ ). These tests provided significant evidence for the presence of sites under positive selection. To compare and estimate the approximation of different models to the data, a modified Akaike’s Information Criterion (AICc) statistics was calculated from likelihood scores, since likelihood scores are always better for more complex models (13) and AICc takes account of sample size in comparison with traditional AIC (14). Among all candidate models, M3 was the best approximation model to the data. M8 had a lower AICc value and was the second-best approximation model overall.

Sites subjected to positive selection may be involved in important function, so we used the Bayes’ theorem to calculate the posterior probabilities of  $\omega$  classes for each site in the site-specific models M2a, M3, and M8. Sites with high probabilities of coming from the class with  $\omega > 1$  are likely to be under positive selection (15). As a result, five codon sites were inferred to be under diversifying positive selection with posterior probabilities  $> 90\%$  by using the Bayes Empirical Bayes (BEB) approach under M2a



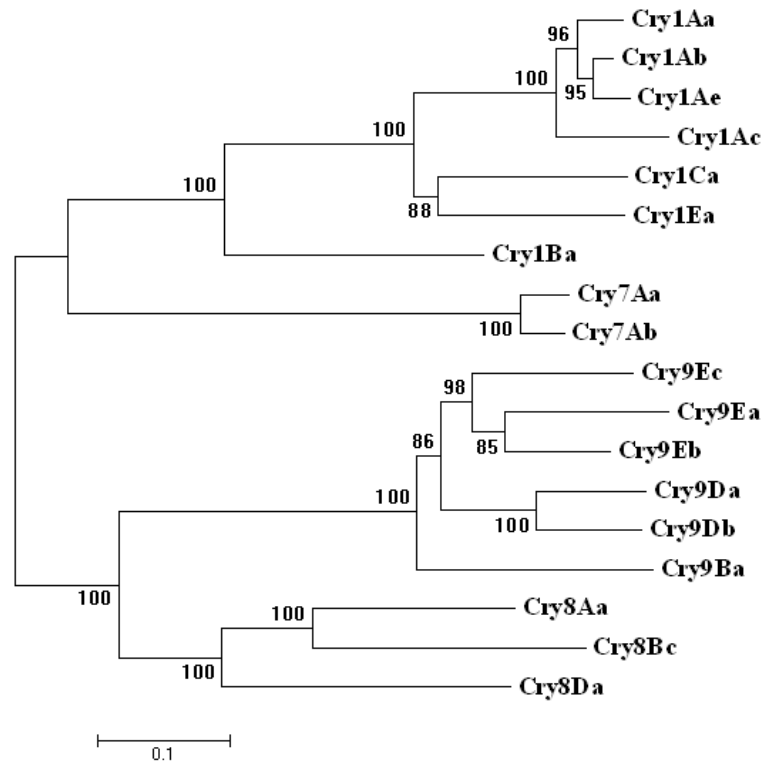
**Fig. 1** Multiple alignment of Domains II and III of 18 Cry proteins. The shading reflects the conservation profile at 60% consensus of amino acids.

(Table 1). The same sites were also identified under the more complex model M8, with 19 extra sites of posterior probabilities > 90% (Table 1). Using the Naive Empirical Bayes (NEB) approach under M3, 24 sites with posterior probabilities > 90% were identified, with one extra site (385) and one missing site (714) when comparing to M8. Sites inferred to be under positive selection in the best approximation model M3 were mapped onto the tertiary structure of Cry1Aa protein (Figure 3). It is clear that all these positive selection sites are located in Domain II (384, 385, 396, 400, 407, 424, 441, 462, 463, 479, 484, 485, 534, 536, 538, 544, and 547) or Domain III (589, 590, 611, 613, 651, and 711). Moreover, most of them are exposed. Domain II has been demonstrated to participate in the receptor recognition and hence determines insect specificity, while Domain III may

also be involved in receptor binding and specificity determination (3). Therefore, we can compare our predictions of the sites potentially under positive selection with the sites identified to be important from the detailed biochemical and genetic characterizations of Cry proteins.

## Discussion

Although insecticidal proteins of *B. thuringiensis* have been widely used in agricultural fields, forest management, and mosquito control for decades, there are several agronomical important insects that are less sensitive to their action. A more serious problem associated with the use of the toxins is the management of resistance development in the target insect pests.



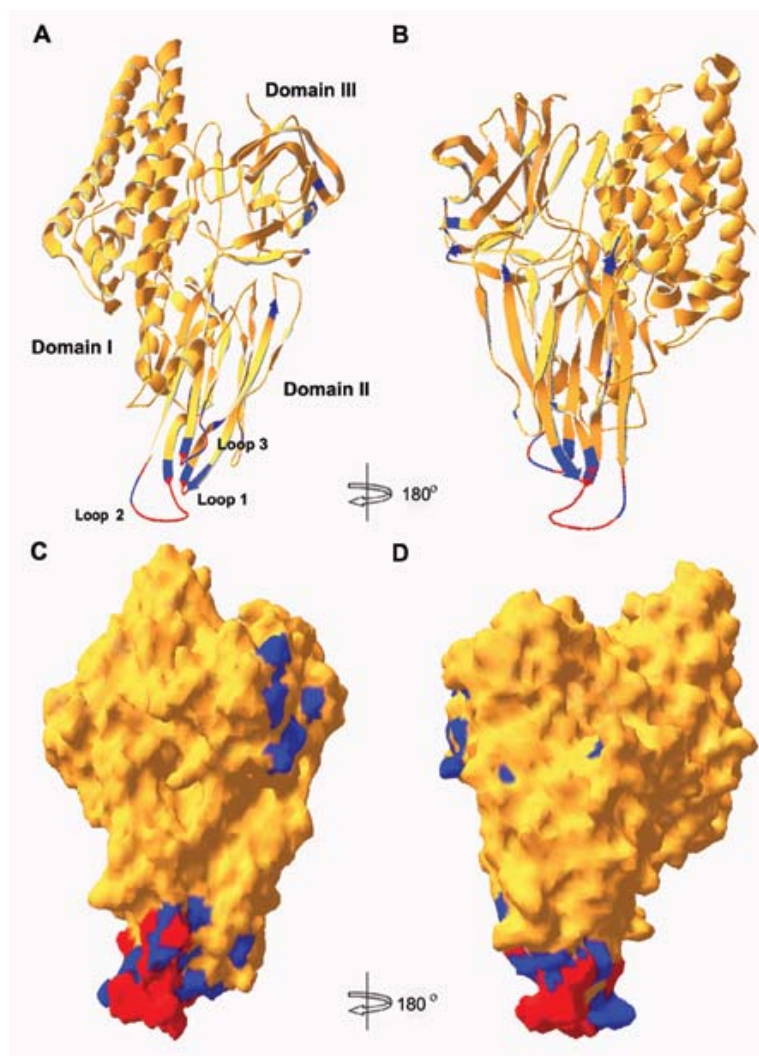
**Fig. 2** Phylogenetic tree of the 18 sequences constructed by the ME method. Numbers under the nodes indicate bootstrap values. Scale bars represent level of amino acid sequence divergence.

**Table 1** Parameter estimates, log likelihood scores, and positively selected sites for the Cry proteins from *B. thuringiensis*\*

Model	Parameter estimate	$\ln L$	$2\Delta \ln L$	AICc	Positively selected sites (probability > 90%)
M0	$\omega_1 = 0.247$	-42,707.52		85,417.29	None
M1a	$p_0 = 0.696, p_1 = 0.304;$ $\omega_0 = 0.128, \omega_1 = 1.000$	-41,394.82		82,794.44	Not allowed
M2a	$p_0 = 0.687, p_1 = 0.276,$ $(p_2 = \mathbf{0.037});$ $\omega_0 = 0.134, \omega_1 = 1.000,$ $\omega_2 = \mathbf{2.798}$	-41,369.92	(M1a vs M2a) 49.80 ( $p < 0.001$ )	82,750.92	400, <b>463, 472</b> , 534, 538
M3	$p_0 = 0.355, p_1 = 0.436,$ $p_2 = 0.209;$ $\omega_0 = 0.0459, \omega_1 = 0.259,$ $\omega_2 = \mathbf{4.962}$	-41,138.80	(M0 vs M3) 3,137.44 ( $p < 0.001$ )	82,292.60	384, 385, 396, 400, 407, 424, 441, 462, <b>463, 472</b> , 479, 484, 485, 534, 536, <b>538</b> , 544, 547, 589, 590, 611, 613, 651, 711
M7	$p = 0.556, q = 1.200$	-41,237.62		82,480.04	Not allowed
M8	$p_0 = 0.874, p = 0.823,$ $q = 2.903, (p_1 = \mathbf{0.125});$ $\omega = \mathbf{1.468}$	-41,146.47	(M7 vs M8) 182.30 ( $p < 0.001$ )	82,304.02	<b>384, 396, 400, 407, 424</b> , 441, <b>462, 463, 472, 479</b> , 484, <b>485</b> , <b>534</b> , 536, <b>538, 544</b> , 547, 589, <b>590</b> , 611, 613, 651, <b>711</b> , 714

\*The maximum likelihood estimates showing the presence of positive selection are in boldface. The parameters  $p$  and  $q$  describe the shape of the beta distribution of  $\omega$ , and  $p_0$ ,  $p_1$ , and  $p_2$  are the proportions of codons belonging to each category. Proportions that are not free parameters are in parentheses. Sites inferred to be under positive selection with probabilities > 99% are in boldface.





**Fig. 3** Mapping of the residues identified to be under positive selection on the tertiary structure of Cry1Aa protein (PDB code: 1ciy). **A** and **B**: Ribbon diagram of Cry1Aa. **C** and **D**: The molecular surface of Cry1Aa. Sites identified to be under positive selection in the best approximation model M3 with posterior probabilities > 90% are shown in blue color. Three loops in Domain II mentioned in the text are shown in red color.

Several strategies are being developed to enhance the efficacy of the toxins, such as developments of hybrid toxins and searches for novel toxin sequences. On the other hand, analysis of the mechanism of toxin/receptor interactions and insect resistance to the toxins can help to develop hybrid toxins or novel toxins with enhanced insecticidal activity or specificity (2, 3).

The accuracy and power of the phylogenetic maximum likelihood analysis to detect positive selection has been extensively tested in that they account for variable selective pressures by assuming a statistical distribution of the  $\omega$  ratio among sites. It has been well demonstrated that this approach is fairly reliable (12, 16, 17). By analyzing DNA sequences from

18 *cry* genes, the LRT analysis comparing the models M1a vs M2a, M0 vs M3, and M7 vs M8 provided significant evidence that nonsynonymous substitution rates varied widely across sites in the evolution of *cry* genes. Among all candidate models, M3 was the best approximation model to the data according to the AICc analysis (14). The sites inferred to be under positive selection by the BEB approach under M8 and by the NEB approach under M3 are almost the same. The only disagreement is that M3 identified one extra site (385) and missed one site (714) when comparing to M8. Though it is suggested that the NEB approach may lead to unreliable posterior probability calculations in small datasets and the BEB approach is more powerful than the NEB approach

(18), both are very powerful for inferring positively selected sites in large datasets. Interestingly, all sites found to be under positive selection in M2a, M3, and M8 are in Domain II or III while no sites were found in Domain I. Moreover, in all the sites identified to be under positive selection, the majority of them are exposed. Surface-exposed loops may be involved in receptor recognition, and therefore selective pressures should lead to its diversification (19).

Domain I of Cry proteins has been shown to play a critical role in membrane insertion and lytic pore formation (5). Phylogenetic analysis on the domain of  $\delta$ -endotoxins reveals that Domain I is the most conserved among all  $\delta$ -endotoxins (2). Most of the mutants in Domain I result in low or no toxicity on the tested insects (20). It is logical to assume that the Cry protein, because of its essential function(s) in the biological system, retains its function intact throughout evolution of organisms despite allowing diversification elsewhere in its structure. Domain II has been considered to play an important role in the function of receptor binding, and the process of receptor binding is mainly contributed to the three surface-exposed loops (3, 6, 7). From Figure 3, we can see that several sites identified as likely to be under positive selection are located in the three surface-exposed loops (site 385 in Loop 1, sites 462, 463, 472, 479 in Loop 2, and sites 534, 536, 538, 544 in Loop 3). Thus, these positively selected sites may be involved in receptor binding and specificity determination. Similarly, Domain III can also function for receptor binding and insect specificity (8, 9). Six positively selected sites (589, 590, 611, 613, 651, and 711) are identified to be located in this domain (Table 1 and Figure 3). Aronson *et al* (21) showed that changes of sites 589 and 590 in Cry1Ac protein could result in a significantly decreased toxicity against *Manduca sexta* but less significant effect on the *Heliothis virescens* larvae, implying that the two sites must be involved in receptor binding. Maagd *et al* (22) and Herrer *et al* (9) both identified that the amino acid substitutions of sites 611 and 613 in Cry1Ca protein could completely eliminate its activity against *Spodoptera exigua* but no obvious elimination against *M. sexta*, revealing the importance of these residues in determining the specificity of Cry proteins. Experimental evidence does show that the site 651 of Cry1Ac protein might be involved in its initial binding to the receptor (23).

Currently, there is a large number of data from mutagenesis studies available for Cry proteins. The

availability of these resources undoubtedly provides an opportunity for assessing the validity and usefulness of our results. As discussed above, among all the 24 positively selected sites, 14 sites present experimental evidence. Moreover, all results of the experimental data suggested that these residues are involved in receptor binding. Likewise, the remaining 10 residues (residues 384, 396, 400, and 407 adjacent to Loop 1, residue 547 adjacent to Loop 3, and residues 424, 441, 484, 485, and 711) were found to be under positive selection in our study. However, there exists little experimental data about these sites. We expect that these sites may also play an important role in specificity determination and would provide targets for further structural and functional analyses on Cry proteins. However, a problem needs to be addressed is that several previous studies using site-directed mutagenesis analysis showed other certain amino acids in Domains II and III involved in specificity determination, whereas we failed to detect them in our experiment. Since the maximum likelihood method itself has been proved to be extremely powerful, we think that this result may be due to the treatment of the data. The *cry* gene family is a large and heterogeneous family, and aligning some Cry proteins is quite difficult due to low sequence homology combined with many insertions and deletions. Therefore, here we just analyzed with some local clades of 18 sequences to keep genetic divergence reasonably small and keep alignments robust for LRT analysis. When we analyzed with other local clades (data not shown), the evidence for positive selection was also addressed. It is suggested that the positive selection does act on the *cry* gene family. Gram-positive spore-forming entomopathogenic bacteria can produce a large variety of protein toxins to help them invade, infect, and finally kill their hosts, through their actions on the insect midgut (24). In this study, the Cry protein we analyzed is one type of such protein toxins, and there is much evidence that certain Cry protein can involve in binding to the midgut brush-border membrane of specific insects (8, 20). We here interpret these results as indicating that the potential pressures driving the diversification of Cry proteins may be in an attempt to adapt for the "arm race" between  $\delta$ -endotoxins and the targeted insects, or to enlarge their target spectra, hence result in the functional divergence among different Cry toxins.

Applying statistical tests of molecular adaptation to nucleotide sequence data can help investigators identify specific amino acid substitutions for fur-

ther experiments. We believe that positive selection site identification in well-sampled datasets, combined with structural and functional information, can provide a valuable framework for identifying and studying the evolution mechanism of Cry proteins. Information about the positively selected sites upon Cry proteins can provide valuable insights into the nature of interactions with their receptors. Our analysis suggests that the high divergence in some regions of the proteins may not result from the lack of functional constraint, but rather from the positive selection promoting rapid evolution to their specified targets. The positively selected sites identified here are candidates to test for the functional importance (receptor binding or specificity determination) of Cry proteins.

## Materials and Methods

### Sequence data and alignment

Most of the *cry* genes are highly divergent, and it is problematic to align such cases with confidence. For these reasons, we just performed our analyses with local clades of 18 sequences (randomly selected from the subfamilies of Cry1, Cry7, Cry8, and Cry9), which could keep genetic divergence reasonably small and keep alignments robust for LRT analysis. However, the evidence of positive Darwinian selection was also addressed when analyzing other local clades (data not shown). The GenBank accession numbers of the 18 sequences are shown in Table 2. The sequences were first aligned at the amino acid level by using ClustalX (25) with the default settings and then were adjusted manually. The corresponding nucleotide sequence alignment was generated according to the protein sequence alignment (<http://bioweb.pasteur.fr/seqanal/interfaces/protal2dna.html>).

### Phylogeny reconstruction and likelihood ratio test

The phylogeny of the Cry protoxin sequences was reconstructed from the alignment of amino acid sequences by using methods of ME, NJ in MEGA3 package (26), and MP in Phylip package (27). The reliability of the tree was evaluated by the bootstrap method with 1,000 replications. To analyze the possibility of positive selection acting on specific codons and to infer amino acid sites under positive selection on the Cry proteins, we applied the maximum likeli-

**Table 2** Accession numbers of the 18 sequences used in this study

Sequence	Accession number
Cry1Ab	AAA22330
Cry1Ac	AAA22331
Cry1Ae	AAA22410
Cry1Ba	CAA29898
Cry1Ca	CAA30396
Cry1Ea	CAA37933
Cry7Aa	AAA22351
Cry7Ab	AAA21120
Cry8Aa	AAA21117
Cry8Bc	CAD57543
Cry8Da	BAC07226
Cry9Ba	CAA52927
Cry9Da	BAA19948
Cry9Db	AAX78439
Cry9Ea	BAA34908
Cry9Eb	CAC50780
Cry9Ec	AAC63366

hood method of Nielsen and Yang (28) implemented in the codeml program from the PAML package (11). Several site-specific models that allow for various  $d_N/d_S$  ratios among sites were used to detect positive selection (18, 29). In the simplest model (one-ratio model, M0), it is assumed that the  $\omega$  ratio is an average over all the sites. The “nearly neutral” model (M1a) allows for conserved sites where  $0 < \omega < 1$  and completely neutral sites where  $\omega = 1$ . The “positive selection” model (M2a) adds a third class to M1a in which  $\omega > 1$ . Model M3 (discrete) has three classes with proportions  $p_0$ ,  $p_1$ , and  $p_2$  as well as  $\omega_0$ ,  $\omega_1$ , and  $\omega_2$  values estimated from the data. Model M7 ( $\beta$ ) assumes a beta distribution over the interval (0, 1) and therefore does not allow for sites with  $\omega > 1$ , providing a flexible null hypothesis for testing positive selection. Model M8 ( $\beta$  and  $\omega$ ) adds an extra class of sites to M7, so that a proportion  $p_0$  of sites comes from the  $\beta$  distribution  $B(p, q)$  and the remaining sites ( $p_1 = 1 - p_0$ ) have a  $\omega$  ratio estimated from the data that can be greater than 1. LRT analyses were conducted to compare M1a with M2a, M0 with M3, and M7 with M8, respectively. The level of significance is calculated as twice the difference of the likelihood scores ( $2\Delta \ln L$ ) estimated by each model, and the null distribution of these results can be approximated by  $\chi^2$  distribution with the number of degrees of freedom calculated as the difference in the number of estimated parameters between models

(15). Only models M2a, M3, and M8 can detect sites under positive selection. The AICc analysis was used to the non-nested models (14). It is indicated that AICc has more advantages than traditional AIC since it considers the sample size that may influence the calculation result of  $d_N$  and  $d_S$  (14, 30). Then, in order to identify which sites potentially belong to the positively selected class identified by M2a and M8, the parameter estimates from these models were used to calculate the posterior probabilities that an amino acid site belongs to a particular  $\omega$  site class by using the BEB approach (18). For M3, the NEB approach was used to identify the sites under positive selection because no BEB approach was implemented. Sites with a high probability of coming from the class with  $\omega > 1$  are likely to be under positive selection. In order to establish the relevance of these findings to the tertiary structure of Cry proteins, we used the tertiary structure of Cry1Aa protein (PDB code: 1ciy) as a working model to plot the positively selected sites with the posterior probabilities  $> 90\%$  in the best approximation model according to the AICc analysis. The plotting procession was performed by the Swiss-PdbViewer (31).

## Acknowledgements

We thank colleagues in Beijing Institute of Genomics, Chinese Academy of Sciences for helpful discussions. This work was supported by the National Natural Science Foundation of China (No. 30571009).

## Authors' contributions

JYW participated in the design of the study, performed the analysis, and drafted the manuscript. FQZ conceived the study and helped to draft the manuscript. JB helped to analyze the data and SQ provided useful suggestions to the manuscript. QYB participated in its design, helped to draft the manuscript, and supervised the whole project. All authors read and approved the final manuscript.

## Competing interests

The authors have declared that no competing interests exist.

## References

1. Höfte, H. and Whiteley, H.R. 1989. Insecticidal crys-

tal proteins of *Bacillus thuringiensis*. *Microbiol. Rev.* 53: 242-255.

2. Bravo, A. 1997. Phylogenetic relationships of *Bacillus thuringiensis* delta-endotoxin family proteins and their functional domains. *J. Bacteriol.* 179: 2793-2801.
3. Schnepf, E., *et al.* 1998. *Bacillus thuringiensis* and its pesticidal crystal proteins. *Microbiol. Mol. Biol. Rev.* 62: 775-806.
4. Crickmore, N., *et al.* 1998. Revision of the nomenclature for the *Bacillus thuringiensis* pesticidal crystal proteins. *Microbiol. Mol. Biol. Rev.* 62: 807-813.
5. Boonserm, P., *et al.* 2005. Crystal structure of the mosquito-larvicidal toxin Cry4Ba and its biological implications. *J. Mol. Biol.* 348: 363-382.
6. Smedley, D.P. and Ellar, D.J. 1996. Mutagenesis of three surface-exposed loops of a *Bacillus thuringiensis* insecticidal toxin reveals residues important for toxicity, receptor recognition and possibly membrane insertion. *Microbiology* 142: 1617-1624.
7. Tuntitippawan, T., *et al.* 2005. Targeted mutagenesis of loop residues in the receptor-binding domain of the *Bacillus thuringiensis* Cry4Ba toxin affects larvicidal activity. *FEMS Microbiol. Lett.* 242: 325-332.
8. de Maagd, R.A., *et al.* 1999. Domain III of the *Bacillus thuringiensis* delta-endotoxin Cry1Ac is involved in binding to *Manduca sexta* brush border membranes and to its purified aminopeptidase N. *Mol. Microbiol.* 31: 463-471.
9. Herrero, S., *et al.* 2004. Mutations in the *Bacillus thuringiensis* Cry1Ca toxin demonstrate the role of domains II and III in specificity towards *Spodoptera exigua* larvae. *J. Biochem.* 384: 507-513.
10. Creevey, C.J. and McInerney, J.O. 2002. An algorithm for detecting directional and non-directional positive selection, neutrality and negative selection in protein coding DNA sequences. *Gene* 300: 43-51.
11. Yang, Z. 1997. PAML: a program for package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* 13: 555-556.
12. Wong, W.S., *et al.* 2004. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics* 168: 1041-1051.
13. Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Autom. Control* 19: 716-723.
14. Zhang, Z., *et al.* 2006. KaKs\_Calculator: calculating Ka and Ks through model selection and model averaging. *Genomics Proteomics Bioinformatics* 4: 259-263.
15. Yang, Z. and Bielawski, J.P. 2000. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* 15: 496-503.
16. Aagaard, J.E. and Phillips, P. 2005. Accuracy and power of the likelihood ratio test for comparing evolutionary rates among genes. *J. Mol. Evol.* 60: 426-433.



17. Yang, Z. 2005. The power of phylogenetic comparison in revealing protein function. *Proc. Natl. Acad. Sci. USA* 102: 3179-3180.
18. Yang, Z., *et al.* 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* 22: 1107-1118.
19. Jiggins, F.M., *et al.* 2002. Host-symbiont conflicts: positive selection on an outer membrane protein of parasitic but not mutualistic Rickettsiaceae. *Mol. Biol. Evol.* 19: 1341-1349.
20. Saraswathy, N. and Kumar, P.A. 2004. Protein engineering of  $\delta$ -endotoxins of *Bacillus thuringiensis*. *Electron. J. Biotechnol.* 7: 178-188.
21. Aronson, A.I., *et al.* 1995. Mutagenesis of specificity and toxicity regions of a *Bacillus thuringiensis* protoxin gene. *J. Bacteriol.* 177: 4059-4065.
22. de Maagd, R.A., *et al.* 1999. Identification of *Bacillus thuringiensis* delta-endotoxin Cry1C domain III amino acid residues involved in insect specificity. *Appl. Environ. Microbiol.* 65: 4369-4374.
23. Lee, M.K., *et al.* 1999. Identification of residues in domain III of *Bacillus thuringiensis* Cry1Ac toxin that affect binding and toxicity. *Appl. Environ. Microbiol.* 65: 4513-4520.
24. de Maagd, R.A., *et al.* 2003. Structure, diversity, and evolution of protein toxins from spore-forming entomopathogenic bacteria. *Annu. Rev. Genet.* 37: 409-433.
25. Thompson, J.D., *et al.* 1997. The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25: 4876-4882.
26. Kumar, S., *et al.* 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinform.* 5: 150-163.
27. Felsenstein, J. 1989. PHYLIP—phylogeny inference package (version 3.2). *Cladistics* 5: 164-166.
28. Nielsen, R. and Yang, Z. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148: 929-936.
29. Yang, Z. and Nielsen, R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* 19: 908-917.
30. Zhang, Z., *et al.* 2006. Computing Ka and Ks with a consideration of unequal transitional substitutions. *BMC Evol. Biol.* 6: 44.
31. Guex, N. and Peitsch, M.C. 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 18: 2714-2723.