Supplementary Tables and Figures

Conservation Genetics

Transcriptome annotation reveals minimal immunogenetic diversity among Wyoming toads, *Anaxyrus baxteri*

Kara B. Carlson[1], Dustin J. Wcisel[1], Hayley D. Ackerman[1], Jessica Romanet[1],
Emily F. Christiansen[2], Jennifer N. Niemuth[2], Christina Williams[1], Matthew Breen[1,3,4],
Michael K. Stoskopf[2,5], Alex Dornburg[6*], and Jeffrey A. Yoder[1,3,4*]

1. Department of Molecular Biomedical Sciences, North Carolina State University, Raleigh, NC, USA
2. Environmental Medicine Consortium, North Carolina State University, Raleigh, NC, USA
3. Comparative Medicine Institute, North Carolina State University, Raleigh, NC, USA
4. Center for Human Health and the Environment, North Carolina State University, Raleigh, NC, USA
5. Department of Clinical Sciences, North Carolina State University, Raleigh, NC, USA
6. Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, Charlotte, NC USA

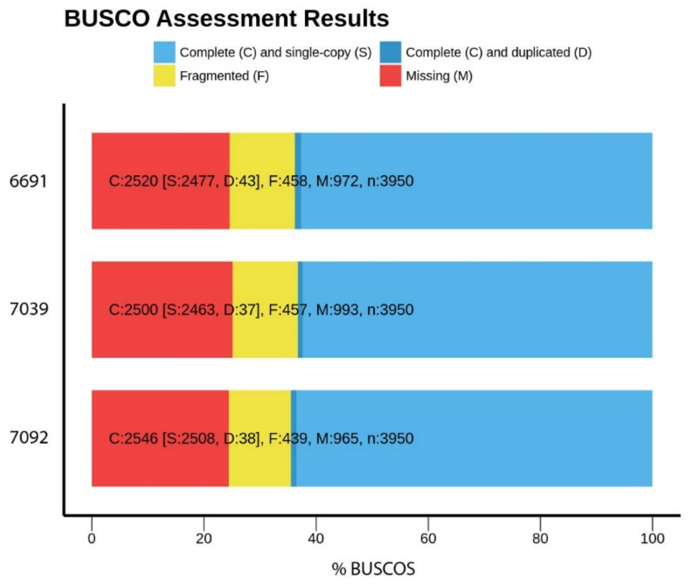
* Corresponding authors at:
Alex Dornburg, E-mail address: adornbur@uncc.edu. Phone: +1 704-687-8437
Jeffrey A. Yoder, E-mail address: jeff_yoder@ncsu.edu. Phone: +1 919-515-7406

Contents:

Supplementary Table S1. Summary of Wyoming toad transcriptome analysis 2
Supplementary Fig. S1. BUSCO assessment results from Wyoming toad transcriptomes. 2
Supplementary Table S2. Duplicated sequences in all three Wyoming toads....3
Supplementary Fig. S2. Summary of annotated transcriptomes based on GO terms 4
Supplementary Table S3. Representative Wyoming toad TLR transcripts and GenBank IDs....5
Supplementary Table S4. Representative Wyoming toad MHC class I transcripts and GenBank IDs 5
Supplementary Fig. S3. Wyoming toad MHC class I UA sequences. 6
Supplementary Fig. S4. Wyoming toad MHC class I UB sequences. 6
Supplementary Fig. S5. Wyoming toad MHC class I UC sequences. 7
Supplementary Fig. S6. Wyoming toad MHC class I UD sequences. 7
Supplementary Table S5. Representative Wyoming toad MHC class II transcripts and GenBank IDs 8
Supplementary Fig. S7. Wyoming toad MHC class II alpha chain sequences. 9
Supplementary Fig. S8. Wyoming toad MHC class II beta chain sequences. 9
Supplementary Table S6. Sequence identity between bufonid TLR ectodomains 10
Supplementary Table S7. Sequence identity between bufonid TLR TIR domains 10

Supplementary Table S1. Summary of Wyoming toad transcriptome analysis

	Toad 6691	Toad 7039	Toad 7092
Total Trinity Genes	245,790	134,581	143,205
Total Trinity Transcripts	332,088	199,248	209,600
Percent GC	45.07	45.31	45.41
Contig N50 (ALL)	1,709	2,041	2,038
Median Contig Length (ALL)	371	426	425
Avg Contig (ALL)	820.4	969.95	961.91
Total Assembled Bases (ALL)	272,445,109	193,259,855	201,616,519
Contig N50 (LONGEST)	709	1,096	1,044
Median Contig Length (LONGEST)	319	337	337
Avg Contig (LONGEST)	555.6	660.94	652.23
Total Assembled Bases (LONGEST)	136,560,916	88,949,732	93,402,820
Total Paired Reads	79,216,573	61,479,313	65,858,371
Overall Alignment Rate	91.63%	90.71%	91.31%

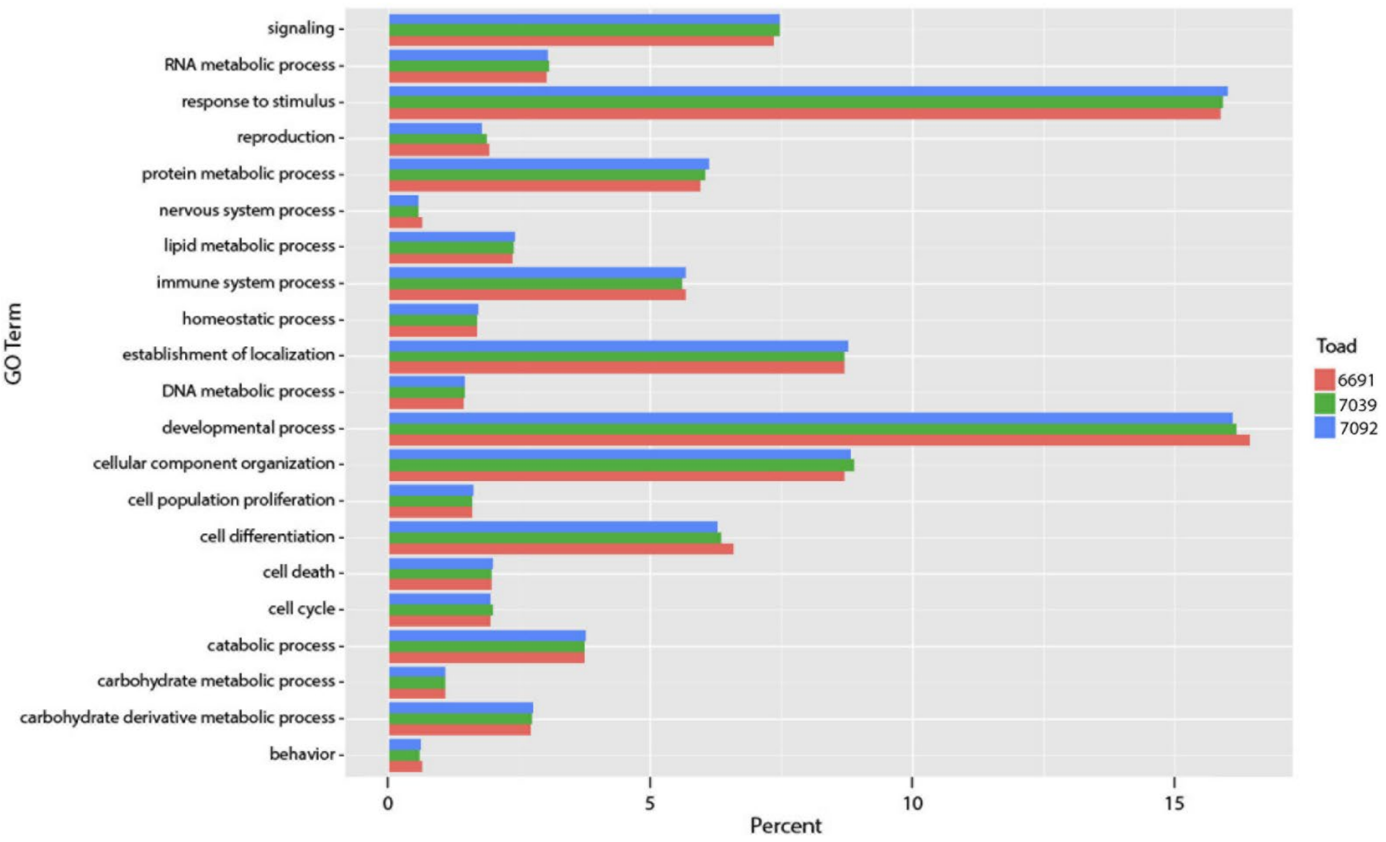


Supplementary Fig. S1. BUSCO assessment results from Wyoming toad transcriptomes. BUSCO tetrapod orthologs were cataloged as complete and single copy, complete and duplicated, fragmented or missing in each of the three toads (6691, 7039 and 7092).

Supplementary Table S2. Duplicated sequences in all three Wyoming toads*

BUSCO ID	OrthoDB Annotation	Human Uniprot ID
EOG090701VI	thyroid peroxidase	P07202
EOG0907020N	DEAH box polypeptide	Q8IX18
EOG090702QL	HBS1-like translational GTPase	D9YZV0
EOG0907036F	peroxisomal biogenesis factor 6	A0A024RD09
EOG090703A7	transketolase	P29401
EOG090704A2	flavin containing monooxygenase 4	P31512
EOG090705SC	cholinergic receptor	P02708
EOG090705ZM	tubby like protein 3	O75386
EOG090706G4	dopamine receptor	P14416
EOG090706LB	serine/threonine kinase	O94768
EOG0907071S	Nucleobindin 2	V9HW75
EOG090708E2	Nephrosis 2	Q9NP85
EOG090708FT	NIPA-like domain containing 4	Q0D2K0
EOG090709BR	potassium channel	Q96T55
EOG09070A0P	Coiled-coil domain containing 3	Q9BQI4

* BUSCO analysis, which identifies ultra-conserved, single copy genes, revealed fifteen duplicate genes in all three Wyoming toads. These genes have been annotated via OrthoDB, however further enrichment analysis is impossible with such a small set of genes.



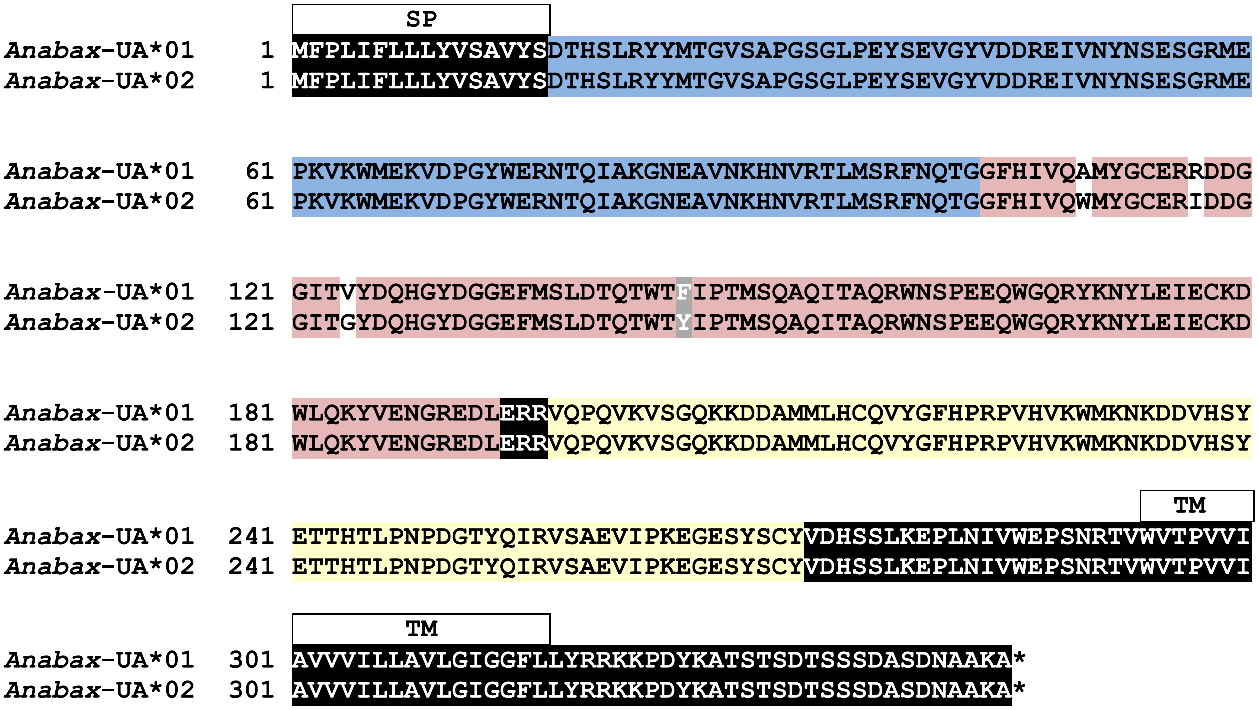
Supplementary Fig. S2. Summary of annotated transcriptomes based on GO terms
Gene ontology (GO) terms from Wyoming toad annotated transcriptomes and each group's percent contribution to total transcriptome. Biological process GO terms were chosen for this analysis as this category includes the terms immune system processes and response to stimulus which include important immune gene annotations. All three toads displayed comparable proportions of GO terms to one another..

Supplementary Table S3. Representative Wyoming toad TLR transcripts and GenBank IDs

Sequence name	Toad 6691	Toad 7039	Toad 7092
TLR1a	GGUS01160622.1	GGUR01153850.1	GGUQ01094842.1
TLR1b	GGUS01148135.1	GGUR01053790.1	GGUQ01193029.1
TLR2	GGUS01236735.1	Not identified	GGUQ01144158.1
TLR3	GGUS01239741.1	GGUR01125592.1	GGUQ01122259.1
TLR4	GGUS01036104.1	GGUR01092862.1	GGUQ01142735.1
TLR5	GGUS01267723.1	GGUR01011083.1	GGUQ01091435.1
TLR8a	GGUS01251716.1	Not identified	GGUQ01056653.1
TLR8b	Not identified	GGUR01052356.1	Not identified
TLR12	GGUS01208850.1	GGUR01100539.1	GGUQ01123333.1
TLR13	GGUS01005708.1	GGUR01162624.1	GGUQ01081784.1
TLR14	GGUS01235967.1	GGUR01048765.1	GGUQ01096496.1
TLR21	GGUS01313948.1	Not identified	GGUQ01185422.1

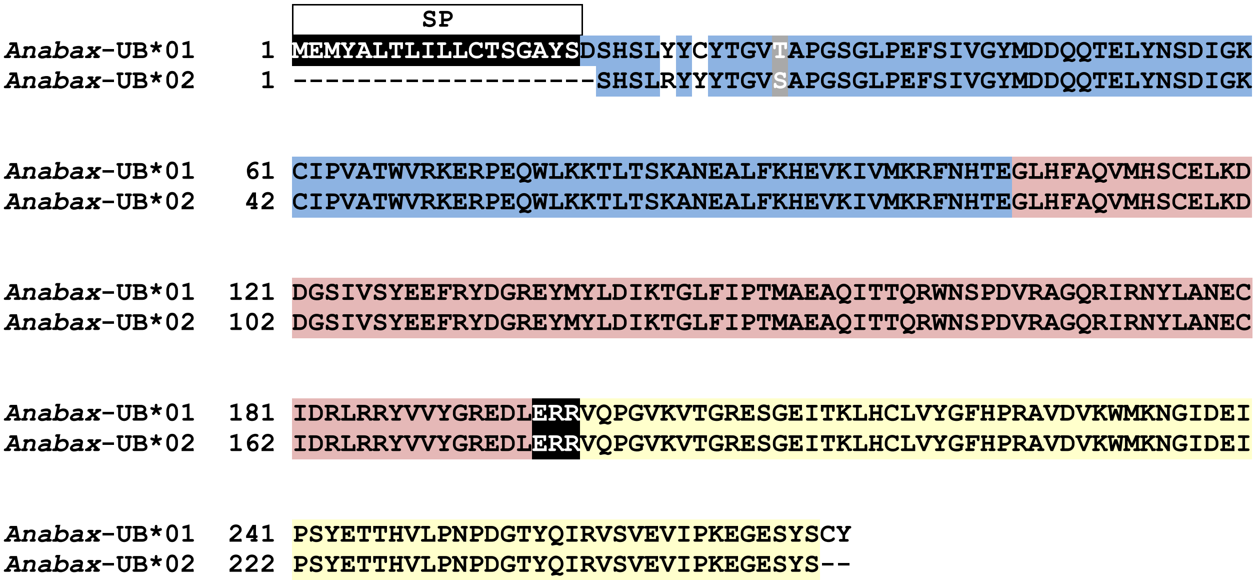
Supplementary Table S4. Representative Wyoming toad MHC class I transcripts and GenBank IDs

Sequence name	Toad 6691	Toad 7039	Toad 7092
*Anabax-UA*01*	Not identified	GGUR01113130.1	Not identified
*Anabax-UA*02*	GGUS01278791.1	Not identified	GGUQ01004097.1
*Anabax-UB*01*	Not identified	GGUR01024835.1	GGUQ01038068.1
*Anabax-UB*02*	GGUS01018260.1	Not identified	Not identified
*Anabax-UC*01*	GGUS01183637.1	Not identified	GGUQ01038171.1
*Anabax-UC*02*	Not identified	GGUR01080688.1	GGUQ01038022.1
*Anabax-UD*01*	GGUS01214789.1	GGUR01025091.1	Not identified
*Anabax-UD*02*	Not identified	Not identified	GGUQ01038222.1
*Anabax-UD*03*	GGUS01191828.1	Not identified	Not identified
β-2-microgobulin	GGUS01148031.1	GGUR01189603.1	GGUQ01034849.1



```
                    SP
Anabax-UA*01     1 MFPLIFLLLYVSAVYSDTHSLRYYMTGVSAPGSGLPEYSEVGYVDDREIVNYNSESGRME
Anabax-UA*02     1 MFPLIFLLLYVSAVYSDTHSLRYYMTGVSAPGSGLPEYSEVGYVDDREIVNYNSESGRME

Anabax-UA*01    61 PKVKWMEKVDPGYWERNTQIAKGNEAVNKHNVRTLMSRFNQTGGFHIVQAMYGCERRDDG
Anabax-UA*02    61 PKVKWMEKVDPGYWERNTQIAKGNEAVNKHNVRTLMSRFNQTGGFHIVQWMYGCERIDDG

Anabax-UA*01   121 GITVYDQHGYDGGEFMSLDTQTWTFIPTMSQAQITAQRWNSPEEQWGQRYKNYLEIECKD
Anabax-UA*02   121 GITGYDQHGYDGGEFMSLDTQTWTYIPTMSQAQITAQRWNSPEEQWGQRYKNYLEIECKD

Anabax-UA*01   181 WLQKYVENGREDLERRVQPQVKVSGQKKDDAMMLHCQVYGFHPRPVHVKWMKNKDDVHSY
Anabax-UA*02   181 WLQKYVENGREDLERRVQPQVKVSGQKKDDAMMLHCQVYGFHPRPVHVKWMKNKDDVHSY
                                                                          TM
Anabax-UA*01   241 ETTHTLPNPDGTYQIRVSAEVIPKEGESYSCYVDHSSLKEPLNIVWEPSNRTVWVTPVVI
Anabax-UA*02   241 ETTHTLPNPDGTYQIRVSAEVIPKEGESYSCYVDHSSLKEPLNIVWEPSNRTVWVTPVVI

                    TM
Anabax-UA*01   301 AVVVILLAVLGIGGFLLYRRKKPDYKATSTSDTSSSDASDNAAKA*
Anabax-UA*02   301 AVVVILLAVLGIGGFLLYRRKKPDYKATSTSDTSSSDASDNAAKA*
```

Supplementary Fig. S3. Wyoming toad MHC class I UA sequences.
The two major forms of *Anabax-UA* were aligned. α1, α2, and α3 domains are color-coded blue, red and yellow, respectively. Signal peptide (SP) and transmembrane (TM) domains are indicated above the alignment. Identical residues are shaded the same color, except for gray which reflects structurally similar residues. An asterisk indicates the presence of a stop codon.

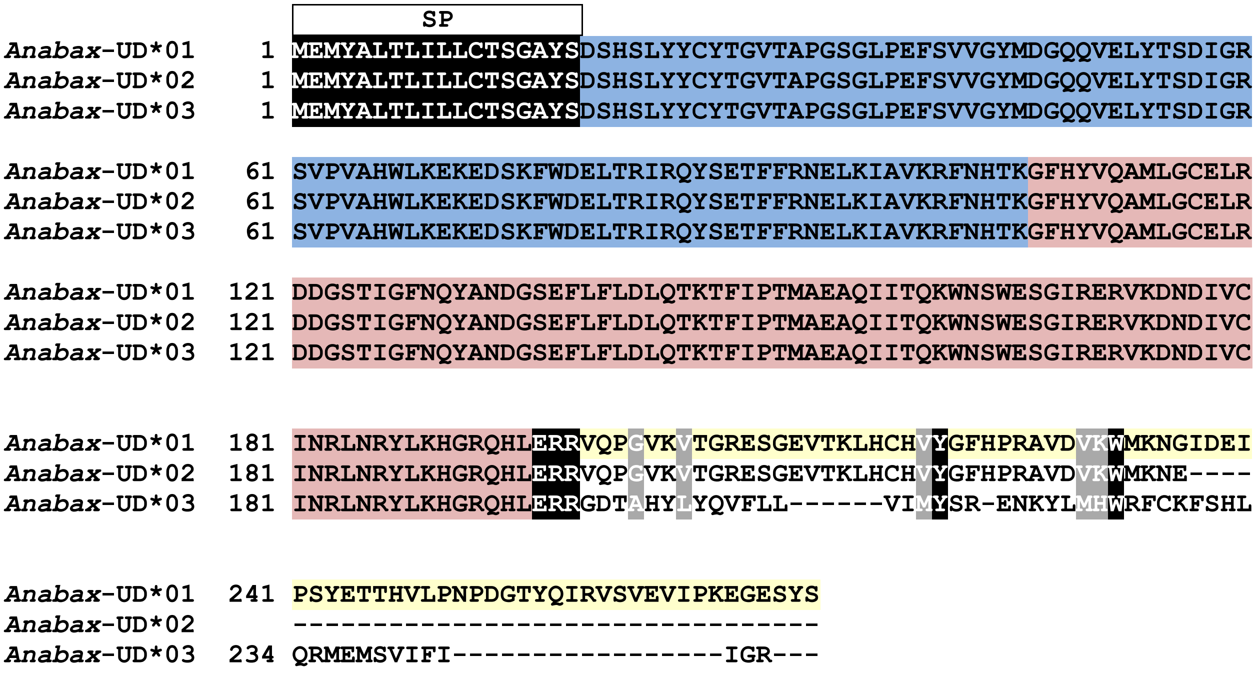


```
                    SP
Anabax-UB*01     1 MEMYALTLILLCTSGAYSDSHSLYYCYTGVTAPGSGLPEFSIVGYMDDQQTELYNSDIGK
Anabax-UB*02     1 -------------------SHSLRYYYTGVSAPGSGLPEFSIVGYMDDQQTELYNSDIGK

Anabax-UB*01    61 CIPVATWVRKERPEQWLKKTLTSKANEALFKHEVKIVMKRFNHTEGLHFAQVMHSCELKD
Anabax-UB*02    42 CIPVATWVRKERPEQWLKKTLTSKANEALFKHEVKIVMKRFNHTEGLHFAQVMHSCELKD

Anabax-UB*01   121 DGSIVSYEEFRYDGREYMYLDIKTGLFIPTMAEAQITTQRWNSPDVRAGQRIRNYLANEC
Anabax-UB*02   102 DGSIVSYEEFRYDGREYMYLDIKTGLFIPTMAEAQITTQRWNSPDVRAGQRIRNYLANEC

Anabax-UB*01   181 IDRLRRYVVYGREDLERRVQPGVKVTGRESGEITKLHCLVYGFHPRAVDVKWMKNGIDEI
Anabax-UB*02   162 IDRLRRYVVYGREDLERRVQPGVKVTGRESGEITKLHCLVYGFHPRAVDVKWMKNGIDEI

Anabax-UB*01   241 PSYETTHVLPNPDGTYQIRVSVEVIPKEGESYSCY
Anabax-UB*02   222 PSYETTHVLPNPDGTYQIRVSVEVIPKEGESYS--
```

Supplementary Fig. S4. Wyoming toad MHC class I UB sequences.
The two major forms of *Anabax-UB* were aligned. α1, α2, and α3 domains are color-coded blue, red and yellow, respectively. Signal peptide (SP) domains are indicated above the alignment. Identical residues are shaded the same color, except for gray which reflects structurally similar residues. Neither of these transcripts encode a stop codon and are likely 3' truncated.

```
                  SP
Anabax-UC*01    1 MMEKMCSLALVLLSLSGAYSDSHSLRYYSIGVSAPGSGLPEFSIIGYVDDQQIELYSGDT
Anabax-UC*02    1 MMEKMCSLALVLLSLSGAYSDSHSLRYYSIGVSAPGSGLPEFSIIGYVDDQQIELYSGDT

Anabax-UC*01   61 GRSVPVAPWLSRNVGLEHWERRTRISKEYEALFKHEVKAAVKRFNHTGGFHFVQVMHNCE
Anabax-UC*02   61 GRSVPVAPWLSRNVGLEHWERRTRISKEYEALFKHEVKAAVKRFNHTGGFHFVQVMHNCE

Anabax-UC*01  121 MRDDGSTTGHQEYRYDGEEYMYLDIKSALFNPTMAEAQIITQRWNSPDIRKGEREKNYLE
Anabax-UC*02  121 MRDDGSTTGHQEYRYDGEEYMYLDIKSALFNPTMAEAQIITQRWNSPDIRKGEREKNYLE

Anabax-UC*01  181 SKCIERLKKYLEFGREDLERRGDFGLSHVYNFATVPH*
Anabax-UC*02  181 SKCIERLKKYLEFGREDLERRVQPGVKVTG-------
```

Supplementary Fig. S5. Wyoming toad MHC class I UC sequences.
The two major forms of *Anabax-UC* were aligned. α1, and α2 domains are color-coded blue and red, respectively. Signal peptide (SP) domains are indicated above the alignment. Identical residues are shaded the same color, except for gray which reflects structurally similar residues. An asterisk indicates the presence of a stop codon.

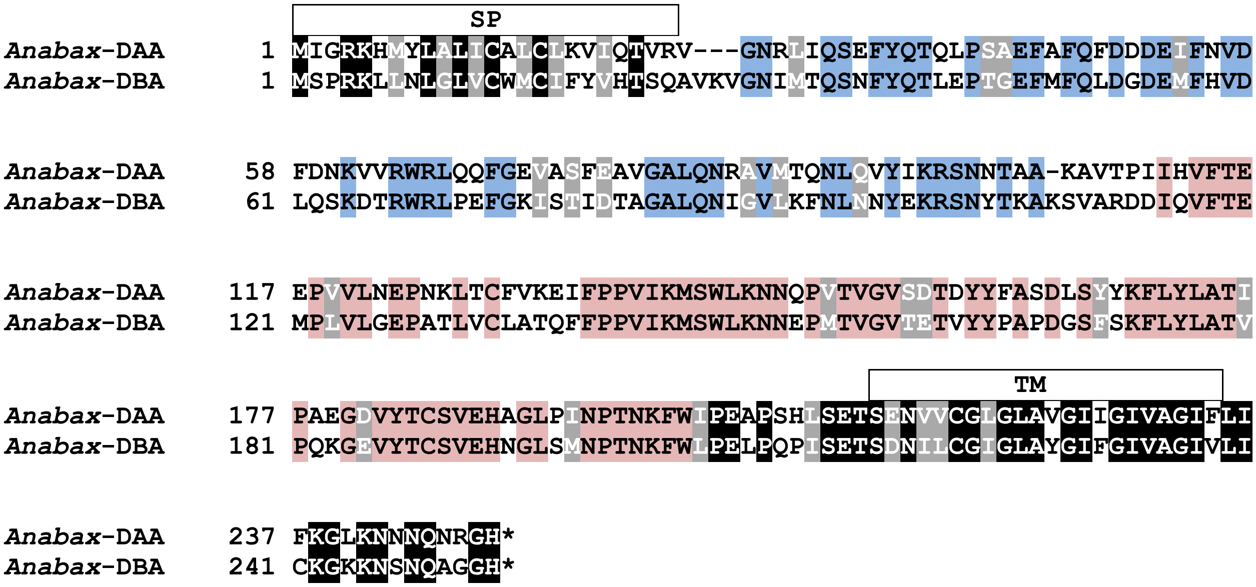


```
                  SP
Anabax-UD*01    1 MEMYALTLILLCTSGAYSDSHSLYYCYTGVTAPGSGLPEFSVVGYMDGQQVELYTSDIGR
Anabax-UD*02    1 MEMYALTLILLCTSGAYSDSHSLYYCYTGVTAPGSGLPEFSVVGYMDGQQVELYTSDIGR
Anabax-UD*03    1 MEMYALTLILLCTSGAYSDSHSLYYCYTGVTAPGSGLPEFSVVGYMDGQQVELYTSDIGR

Anabax-UD*01   61 SVPVAHWLKEKEDSKFWDELTRIRQYSETFFRNELKIAVKRFNHTKGFHYVQAMLGCELR
Anabax-UD*02   61 SVPVAHWLKEKEDSKFWDELTRIRQYSETFFRNELKIAVKRFNHTKGFHYVQAMLGCELR
Anabax-UD*03   61 SVPVAHWLKEKEDSKFWDELTRIRQYSETFFRNELKIAVKRFNHTKGFHYVQAMLGCELR

Anabax-UD*01  121 DDGSTIGFNQYANDGSEFLFLDLQTKTFIPTMAEAQIITQKWNSWESGIRERVKDNDIVC
Anabax-UD*02  121 DDGSTIGFNQYANDGSEFLFLDLQTKTFIPTMAEAQIITQKWNSWESGIRERVKDNDIVC
Anabax-UD*03  121 DDGSTIGFNQYANDGSEFLFLDLQTKTFIPTMAEAQIITQKWNSWESGIRERVKDNDIVC

Anabax-UD*01  181 INRLNRYLKHGRQHLERRVQPGVKVTGRESGEVTKLHCHVYGFHPRAVDVKWMKNGIDEI
Anabax-UD*02  181 INRLNRYLKHGRQHLERRVQPGVKVTGRESGEVTKLHCHVYGFHPRAVDVKWMKNE----
Anabax-UD*03  181 INRLNRYLKHGRQHLERRGDTAHYLYQVFLL------VIMYSR-ENKYLMHWRFCKFSHL

Anabax-UD*01  241 PSYETTHVLPNPDGTYQIRVSVEVIPKEGESYS
Anabax-UD*02      ---------------------------------
Anabax-UD*03  234 QRMEMSVIFI-----------------IGR---
```

Supplementary Fig. S6. Wyoming toad MHC class I UD sequences.
The three major forms of *Anabax-UC* were aligned. α1, α2, and α3 domains are color-coded blue, red and yellow, respectively. Signal peptide (SP) domains are indicated above the alignment. Identical residues are shaded the same color, except for gray which reflects structurally similar residues. None of these transcripts encode a stop codon and are likely 3’ truncated.

Supplementary Table S5. Representative Wyoming toad MHC class II transcripts and GenBank IDs

Sequence name	Toad 6691	Toad 7039	Toad 7092
Anabax-DAA	GGUS01080349.1	GGUR01117947.1	GGUQ01148610.1
Anabax-DBA	GGUS01104323.1	GGUR01016231.1	GGUQ01047262.1
*Anabax-DAB*01*	GGUS01279675.1	GGUR01053143.1	GGUQ01049761.1
*Anabax-DAB*02*	Not identified	GGUR01053272.1	GGUQ01049863.1
Truncated novel class II β1 domain	GGUS01258414.1	GGUR01110179.1	GGUQ01029310.1
Truncated novel class II β1 domain	GGUS01090892.1	GGUR01054024.1	GGUQ01039448.1



```
                          SP
Anabax-DAA      1 MIGRKHMYLALICALCLKVIQTVRV---GNRLIQSEFYQTQLPSAEFAFQFDDDEIFNVD
Anabax-DBA      1 MSPRKLLNLGLVCWMCIFYVHTSQAVKVGNIMTQSNFYQTLEPTGEFMFQLDGDEMFHVD

Anabax-DAA     58 FDNKVVRWRLQQFGEVASFEAVGALQNRAVMTQNLQVYIKRSNNTAA-KAVTPIIHVFTE
Anabax-DBA     61 LQSKDTRWRLPEFGKISTIDTAGALQNIGVLKFNLNNYEKRSNYTKAKSVARDDIQVFTE

Anabax-DAA    117 EPVVLNEPNKLTCFVKEIFPPVIKMSWLKNNQPVTVGVSDTDYYFASDLSYYKFLYLATI
Anabax-DBA    121 MPLVLGEPATLVCLATQFFPPVIKMSWLKNNEPMTVGVTETVYYPAPDGSFSKFLYLATV

                                                   TM
Anabax-DAA    177 PAEGDVYTCSVEHAGLPINPTNKFWIPEAPSHLSETSENVVCGLGLAVGIIGIVAGIFLI
Anabax-DBA    181 PQKGEVYTCSVEHNGLSMNPTNKFWLPELPQPISETSDNILCGIGLAYGIFGIVAGIVLI

Anabax-DAA    237 FKGLKNNNQNRGH*
Anabax-DBA    241 CKGKKNSNQAGGH*
```

Supplementary Fig. S7. Wyoming toad MHC class II alpha chain sequences.
Wyoming toad DAA and DBA were aligned. Identical residues within the α1 and α2 domains are color-coded blue and red, respectively. Signal peptide (SP) and transmembrane (TM) domains are indicated above the alignment. Identical residues are shaded the same color, except for gray which reflects structurally similar residues. An asterisk indicates the presence of a stop codon.

```
                             SP
Anabax-DAB*01    1 MGMLYCISFIFFILLIRLDFCRLSAAVDYMTESKFECHYLNGTQRVRYLHRVFYNQEEIV
Anabax-DAB*02    1 MGMLYCISFIFFILLIRLDFCRLSAAVDYMTESKFECHYLNGTQRVRYLHRVFYNQEEIV

Anabax-DAB*01   61 YFDSNKGYYIPKTEFGKPDADYWNKDKDLIEDRKSSVETFCKHNYGVWKAGAIERKVEPE
Anabax-DAB*02   61 YFDSNKGYYIPKTEFGKPDADYWNKDKDLIEDRKSSVETFCKHNYGVWKAGAIERKVEPE

Anabax-DAB*01  121 IVVALMPNHEEPSTVHHILQCNVFGFYPSEVEVKWYRNGQEETELVTSSALFQNGDWTYR
Anabax-DAB*02  121 IVVALMPNHEEPSTVHHILQCNVFGFYPSEVEVKWYRNGQEETELVQSTEPYHNGDWSYQ

                                                                   TM
Anabax-DAB*01  181 ILVMLETEIQKGDTFTCEVHHSSLKAPHRVDWHPQMSESAKSKRATGIGGIVLGTAFLIV
Anabax-DAB*02  181 ILVMLETEIQKGDTFTCEVHHSSLKAPHRVDWHPQSSDAAKNKMATGIVGFVLGVVFFIV

                    TM
Anabax-DAB*01  241 GIVIYMRGRKGQTPFRVPQSDHFPHT*
Anabax-DAB*02  241 GIVIYMRARKGQMSFPGQQSERFLPQ*
```

Supplementary Fig. S8. Wyoming toad MHC class II beta chain sequences.
Wyoming toad DAB*01 and DAB*02 were aligned. Identical residues within the β1 and β2 domains are color-coded blue and red, respectively. Signal peptide (SP) and transmembrane (TM) domains are indicated above the alignment. Identical residues are shaded the same color, except for gray which reflects structurally similar residues. An asterisk indicates the presence of a stop codon.

Supplementary Table S6. Sequence identity between bufonid TLR ectodomains*

	TLR8a, CMT	TLR8a, CNT	TLR8b, WYT	TLR8a, WYT	TLR8b, CMT	TLR8b, CNT
TLR8a, CMT		92.87	52.26	52.26	52.73	52.26
TLR8a, CNT	92.87		52.24	52.24	53.44	52.34
TLR8b, WYT	52.26	52.24		100	88.8	89.49
TLR8a, WYT	52.26	52.24	100		88.8	89.49
TLR8b, CMT	52.73	53.44	88.8	88.8		90.21
TLR8b, CNT	52.26	52.34	89.49	89.49	90.21	

* GenBank sequence identifiers for Wyoming toad (WYT) TLRs are in **Supplemental Table S3** and sequence identifiers for common toad (CMT) and cane toad (CNT) TLRs are included in **Figure 2**.

Supplementary Table S7. Sequence identity between bufonid TLR TIR domains*

	TLR8a, WYT	TLR8a, CMT	TLR8a, CNT	TLR8b, CNT	TLR8b, WYT	TLR8b, CMT
TLR8a, WYT		96.05	97.37	87.5	88.82	88.16
TLR8a, CMT	96.05		98.68	87.5	87.5	86.84
TLR8a, CNT	97.37	98.68		88.82	88.82	88.16
TLR8b, CNT	87.5	87.5	88.82		97.37	96.71
TLR8b, WYT	88.82	87.5	88.82	97.37		97.37
TLR8b, CMT	88.16	86.84	88.16	96.71	97.37	

* GenBank sequence identifiers for Wyoming toad (WYT) TLRs are in **Supplemental Table S3** and sequence identifiers for common toad (CMT) and cane toad (CNT) TLRs are included in **Figure 2**.