

SUPPLEMENTARY MATERIAL for Dwulit-Smith *et al.*

acl Actinobacteria Assemble a Functional Actinorhodopsin with Natively-Synthesized Retinal

TABLE S1 Metadata of *acl* genomes referred to in this study

Classification	Genome Name	IMG Taxon ID	Recovered Genome			Completeness (%) ⁱ	Contamination (%) ⁱ
			Size (MB)	Contigs	Genes		
acl-A	AAA024-D14 ^a	2264265190	0.78	82	892	58.4	0.98
	AAA023-J06 ^a	2236661001	0.70	98	818	43.7	0.00
	MMS-IA-53 ^b	2775506795	1.33	1	1386	100	0.00
	AAA028-I14 ^c	2545555832	0.78	54	848	50.8	0.00
	AAA028-E20 ^c	2602042079	0.88	51	953	60.2	0.00
	AAA278-O22 ^a	2236661007	1.14	43	1238	78.2	0.42
acl-B	AAA028-A23 ^a	2236661004	0.83	64	913	67.7	0.19
	AAA027-L06 ^d	2505679121	1.16	75	1282	87.4	4.37
	MMS-21-122 ^b	2775506799	1.24	1	1281	100	0.00
	AAA027-J17 ^a	2236661002	0.97	81	1094	73.1	0.84
	AAA278-I18 ^a	2236661006	0.94	54	1037	69.0	0.17
	AAA044-D11 ^c	2524023189	1.15	30	1214	75.2	0.84
acl-C	IMCC26077 ^e	2602042021	1.52	1	1572	100.	0.00
	ME885 ^c	2582580608	0.87	78	906	55.7	4.76
	ME578 ^c	2556921504	0.36	32	390	25.9	3.45
	BIN_10 ^g	NA ^h	0.98	95	1171	75.7	0.00
	ME3864 ^c	2582580572	0.79	70	839	53.0	3.78

^a Ghylin, T. W., Garcia, S. L., Moya, F., Oyserman, B. O., Schwientek, P., Forest, K. T., ... McMahon, K. D. (2014). Comparative single-cell genomics reveals potential ecological niches for the freshwater *acl* Actinobacteria lineage. *The ISME Journal*, 8(12), 2503–2516. These genomes are single-cell amplified genomes (SAGs) recovered from one of three lakes including Lake Mendota, using fluorescence activated cell sorting, subsequent multiple displacement amplification (MDA) followed by sequencing and assembly. Additional details are available in the original publication of Ghylin *et al* 2014.

^b Neuenschwander S.M., Ghai R., Pernthaler J., Salcher M. M. (2018) Microdiversification in genome-streamlined ubiquitous freshwater Actinobacteria. *The ISME Journal*, 12(1), 185-198. This genome is a MAG derived a dilution culture enriched in *acl*-B cells. The DNA was obtained using MDA products that were sequenced as a metagenome, assembled, and binned manually with Sanger sequencing to close gaps. Additional details are available in the original publication of Neuenschwander *et al* 2018.

^c Hamilton J. J., Garcia S. L., Brown B. S., Oyserman B. O., Moya-Flores F., Bertilsson S., Malmstrom R. R., Forest K. T., McMahon K. D. 2017. Metabolic network analysis and metatranscriptomics reveal auxotrophies and nutrient sources of the cosmopolitan freshwater microbial lineage *acl*. *mSystems* 2:e00091-17. Note that the names in the IMG record are actually MEint.metabat.885, MEint.metabat.578, and MEint.3864, respectively. Briefly, they are MAGs recovered from combined assemblies (i.e. bioinformatically pooled metagenomes from 94 samples collected from the top 12 m of Lake Mendota at the deep hole). MAGs were binned using differential coverage based on metagenomes mapped to the combined assembly, and based on k-mer frequencies. Methodological details can be found in Hamilton *et al* 2017 and S. Roux, L-K.Chan, R. Egan, R. R. Malmstrom, K. D. McMahon, M. B. Sullivan (2017) "Ecogenomics of freshwater virophages and their giant virus hosts assessed through time-series metagenomics". *Nature Communications*. 8: article 858.

^d Garcia, S. L., McMahon, K. D., Martinez-Garcia, M., Srivastava, A., Sczyrba, A., Stepanauskas, R., ... Warnecke, F. (2013). Metabolic potential of a single cell belonging to one of the most abundant lineages in freshwater bacterioplankton. *The ISME Journal*, 7(1), 137–147. This genome was a SAG recovered from Lake Stechlin in Germany, using fluorescence activated cell sorting, subsequent multiple displacement amplification (MDA) followed by sequencing and assembly. Additional details are available in the original publication of Garcia *et al* 2013.

^e Kang I., Kim S., Islam M. R., Cho J-C. (2017). The first complete genome sequences of the *acl* lineage, the most abundant freshwater Actinobacteria, obtained by whole-genome-amplification of dilution-to-extinction cultures. *Sci Rep* 7. This genome is a MAG derived from a dilution culture enriched in *acl*-C cells. The DNA was obtained using MDA products that were sequenced as a metagenome, assembled, and binned manually with Sanger sequencing to close gaps. Additional details are available in the original publication of Kang et al 2017.

^g Tsementzi, D., Poretsky, R. S., Rodriguez-R, L. M., Luo, C., & Konstantinidis, K. T. (2014). Evaluation of metatranscriptomic protocols and application to the study of freshwater microbial communities. *Environmental Microbiology Reports*, 6(6), 640–655.

^h Data are available at <http://enve-omics.ce.gatech.edu/data/MTR>

ⁱ Inferred using CheckM and the set of 204 Actinobacteria-specific marker genes, or set to 100% and 0% because genome is listed as complete in the original publication.

SUPPLEMENTARY FIGURE LEGENDS

FIG S1 Xantho-opsin phylogeny groupings provide clear evidence for a distinct grouping of action-opsins. A conserved proline in $\alpha 4$ differentiates actino-opsins (ActRs) from other xantho-opsins (XRs) in the supergroup. Bolded opsins are aligned in Fig. 1A.

FIG S2 Alignments for lycopene cyclase chains (*acl*-CrtYc and *acl*-CrtYd). Proteins from *acl* genome AAA278-O22 are aligned to CrtYc (A) and CrtYd (B) from *Myxococcus xanthus* (Mx), which form the functionally characterized heterodimeric β -cyclase. In this and following five figures, identical (*), highly similar (:), and slightly similar (.) residues are indicated, and transmembrane helix predictions are highlighted in gray. The catalytic PxE(E/D) motif is highlighted in orange.

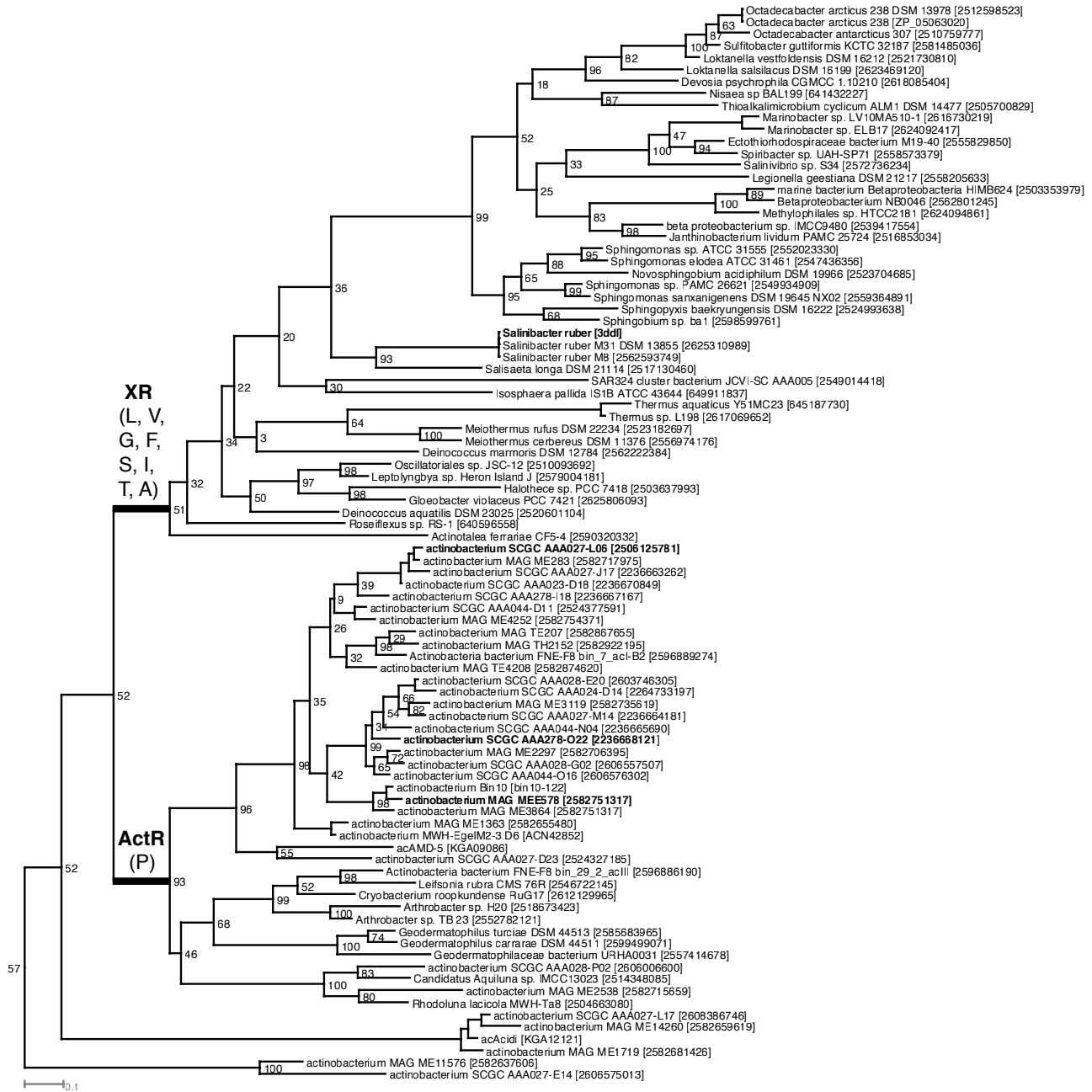
FIG S3 Alignment for 15-15'- β -carotene dioxygenase (*acl*-Blh). The protein from *acl* genome AAA278-O22 is aligned to a functionally characterized 15-15'- β -carotene dioxygenase from Uncultured Marine Bacterium 66A03 (Umb66A03). The histidines indicated for non-heme iron binding are highlighted in orange.

FIG S4 Alignment for predicted spheroidene mono-oxygenase (*acl*-CrtA). The protein from *acl* genome AAA278-O22 is aligned to functionally characterized CrtA proteins from *Rhodobacter sphaeroides* (Rs) and Marine bacterium P99-3 (MbP99-3).

FIG S5 Alignment for predicted γ -carotene 1'-hydroxylase (*acl*-CruF). The protein from *acl* genome AAA278-O22 is aligned to functionally characterized CruF proteins from *Deinococcus radiodurans* R1 (Dr) and *Deinococcus geothermalis* DSM11300 (Ds).

FIG S6 Alignment for glycosyltransferase (*acl*-CruG). The protein from *acl* genome AAA278-O22 is aligned to functionally characterized CruG proteins from *Gemmatimonas aurantiaca* (Ga) and *Synechococcus* sp. PCC7002 (SP).

FIG S7 Alignment for predicted γ carotene 3',4' desaturase (*acl*-CrtD). The protein from *acl* genome AAA278-O22 is aligned to functionally characterized CrtD proteins from Marine bacterium P99-3 (MbP99-3) and *Nonlabens dokdonensis* DSW-6 (Nd).



A

AAA278-O22	-----MGMLIFT-LCGSGWLEIVLKTGVLRRRLKRAALSILPISIFFLIWDAYAIAKGHW	53
Mx	MTYARFLGLFVVVPIFLAWRYR--RTFTARSLAPMGLLLIVVYAATSPWDNLAVKWGLW	58
	:*:::.. : . * : * . * * . * :: : ** *: * *	
AAA278-O22	FFDRQOMLGIIGPFNI PLEEY LFFIVVPLAAILTIEG-----VTTVKPHWRKGEFG	104
Mx	GFDPERIWGIKLG-YL PLEEY LFFALQTLVGLWQAARLARALAPDAQASRPAAETGERR	117
	** ::: ** :***** : * . * :. . : : * ..**	
AAA278-O22	E-----	105
Mx	EGALTAREVAP	128
	*	

B

AAA278-O22	-----MIYSDIAIAAFGISVMVDLFIFKNSMLTRAAFWTSYAIILPFQLLTNWW	49
Mx	MMETKWAYLIHLLGWTLPVIAFQLVVLVRHYKERSGAVLKA-----VLPPAFIMGLY	52
	:: : : ** : *:* : .. : :* : ** :: . :	
AAA278-O22	-----LTSRNIVMYNPDAIIGVRCSA PVEDLLFGF-ALVLSVMGMWEFWGRRGLQRH	101
Mx	LSIADHLAISTGIWNFGEGKHLGVYLGVP PLEEVVFFLITSVLVSLGLALFTGLVALLGE	112
	* . * :. . :** : .***** : : ** :*: * * . * .	
AAA278-O22	----	101
Mx	ARAS	116

Umb66A03	-----MGLMLIDWCALE--LVVFIGLPHGALD	25
AAA278-O22	MEMAKLKTF SRVRTFSSAIVAVAIALSIVFSSWLGVD SLNWQVVM AVVALAIGIPHGALD	60
	:* ::* .: .: ** :*****	
Umb66A03	AAISFSMISSAKRIARLAGILLIYLLLATAFFLIWYQ--LPAFSLIFLLISIIHFGMAD	83
AAA278-O22	HLVTLPKAQPLKM---AIFIAIY--VAIALIAIWAILQWNVWGFIAVVIMSATHFGIGD	114
	::: . * * : * * : * * : * * . : : : : * * * * : *	
Umb66A03	FNASPSKL-----KWP HIIAHGGVVT VWLPLIQKNEVTKLFSILTNGPTPILWDI	133
AAA278-O22	SAFISELNRLKGIQSHLPIWAYA-PAAGALPVVIPLVNSRSTDALQKV---NSELINWHH	170
	. * : . * : * : * * : * * : * * . * . * *	
Umb66A03	LLIFFL--CWSIGVCLHTYETLRSKHYNIAFELIGLIFLAWYAPPLVTFATYFCFIHSRR	191
AAA278-O22	GYTSELQI AVAVIATLSAMTLLSKKRYRDL LDVALLAALASVAPPLVAFVYFGCWHAMR	230
	* . : : . * : * . * : * : * * * * * * * * * * * * * * *	
Umb66A03	HFSFVWKQLQHMSSKMMIGSA-----IILSCTSWLIGGGIYFFLNSKMI---ASEAAL	242
AAA278-O22	HTARLSSLLPRSLA-AYEAGNSWQAFRSAVIPGLPALIGTLIFVALLAGF SHNNVSDSFL	289
	* : : . * : : * : * : * * * * * * * * * * * * * * *	
Umb66A03	QTVFIGLAALTVPHMILIDFIFRPHSSRIKIKN-----	275
AAA278-O22	WLTIVTIWALTVPHMVTA KLDR-----AALKNKSHLR	322
	. : : * * * * * * : : * * * * * * * * * * * * * * *	

AAA278-O22	-----MAQSLTSLIFIIISLINFVTIRIPK----RDEEIKKSVTVL	39
Ga	MIDAVPWSSLGVSF-LVGQAL---CLMVLVSRAPGRSRRPPVSPRLAPRDD-TTVSVI	55
SP	-MDFSPVWIGGLCLFGLLIQGS---GALVVLSRLMKGAVRRSPLTPQASNSDNLAAVTVV	56
	: * . : : : * * * : : * : * : *	
AAA278-O22	APMRNEAENVPEFISALSSQMGVKNLNFVIINDGSTDKTAELLTSVIDGDPREFSFIDSPI	99
Ga	VATLNEAHRIGPCLDGLLQQ-PAPLLEVLVVDSSRSRDGTPELVQTYADRDPRIRLITDDP	114
SP	VPSLNEVERIQPCLDGLSQQ-SYEVREILVVDNSNSTDGTREKVLAKAATDPRFRLLTDDP	115
	. ** . . . : : . * . * : : : : . * * * * : : * * : : .	
AAA278-O22	QRDGLWLGKVSALQSGYES--ARSEFIITLDADVRLQPNAIMRAISQLERLKLDFVSPYPR	157
Ga	LPPGVWVKVWALQTGLQA--ARGAWVLGIDADTVPAAGLVGGVIDAAERDRFDVVSFSR	172
SP	LPQGWVGRPWALNWGFHSSPDSEWILGVDADTQPQRGMIAVSVLQAAEEEGFDLVSLSPQ	175
	** : * : * : * : : : * * * . : : . * . : * . * * * :	
AAA278-O22	QIAQTFAEKLIQPLLHWSWMSTVILRLAEKFPRRSTAVANGQFFVARKNALDAINGFESV	217
Ga	FAGQTAAERLVQPAMLVTLVYRTG-AAGAEQ-QPDRVLANGQCFLARRAVLEQHGGYAVA	230
SP	FILRSPGEWWLQPALLMTLLFRFD-VAGIRQPDQDSVMANGQCFLCRRKVLNVGGYRSA	234
	: : . * : * * : : : . . . : * * * * * : * : * : * : .	
AAA278-O22	STQILDDIELARSLISAGYRGVVTEGSGIASTRMYSSFDEIRQGYGKSLWKAFGGSIGT-	276
Ga	RGSFSDDVTLARHLAMHGARVGFLDGSRIIEVRAYATLREMREWGRSFDLKDSASRVQG	290
SP	AGSFCDDVTLARNIAKAGYRVGFLDGAKIIKVRMYEGMRETWDEWGRSLDLKDATSRTEL	294
	. : * * : * * : * * . : * : * . * * : * : * * : * . *	
AAA278-O22	VIAIAFLFATGILPVLMI LN-----GYLIGWLIYLYIV-FSR--EISA	316
Ga	WLDVVLVWLQALPLPVLLGGVLLWFGAPAALSPYPQWLLVALIATNSFAVLLRILMLWV	350
SP	WGQLWLLTMVQGLPIPLTM---LLFGGIEEGLS---NPFLSGWLDLNVFLLLVRFGMLLA	348
	: : : . * * : : : * . : : : * : .	
AAA278-O22	IRS-----RSNPLFAFLHPLSSALLIYLIISWRNRGTIQWKGRTV-----	357
Ga	LRTSYHE---RGVTYWLSWLSD---AAAARLMTS-MARTPRSWRGRRYASERAPV	399
SP	ISGSYDRQFSVGKSAWLFWLSPFADPFAVARIFLS-ARQKPKAWRGRVYS-----	397
	: : : . : : * : : * : : * : * *	

AAA278-O22	MTGQVGKVKNPEKIAVIGAGIGGLCTAARLAKAGHSVTIFEASSRTGGKCRTEWIGRYAF	60
MbP99-3	-----MKKAIIGSGIAGLAAALRLKKKGYSVFEKNDYAGGKLHAIELGGYRF	50
Nd	-----MKNIAIVIGAGIGGLAAALRLRHQGYSVTIFEKNDYAGGKLHAIKDG YRF	50
	:: : ** : ** : * ** : * : ** : ** .. : ** : : . * *	
AAA278-O22	DTGPSLLTLPVYRDFQRTGDMGRVLELEEVPNSFDYRFHDGTSVKFANLSRKQTLAS	120
MbP99-3	DLGPSLFTLPHLIDELHQLFPDVEID-FNYIKKKTACHYFWEDGTSFEAPADLENFAVKA	109
Nd	DLGPSLFTLPHLVENLFALFPEEIID-FGYKSKAISFHYFWDDGTLFKASTDSSQFIEDA	109
	* ** : ** : : : : : : : * : ** : : : : :	
AAA278-O22	IEQTLGAAAASEWNRVMLRAEAMWDVSREPFIESELKSPLSLLKRPRLIRDLKIISPWK-	179
MbP99-3	--AEIFDEKQNTLSKYLQNSKMYESTKSLFLEKSLHKSNTYFSKQTLKAI--LKIPFLG	165
Nd	--SKVFKEEKSTIKKYLAKSKSKYELTKSLFLEKSLHKATTYFSLDTPVKAI--VHAPFLG	165
	: . . : : : : : : * : * . * . : : : : : : * :	
AAA278-O22	-----SLREIGITNPYLAKIMDRYATYSGSDPRYAPAVLSTIAFVEEAFGAWHIKGGIGT	234
MbP99-3	INNTLDQENKFKSNPKLNQLFNRYATYNGSSPYLTPGIMSMIPYLELGLGTYYPQGGMHR	225
Nd	LNNTLNDENSKFKNPKLQLFNRYATYNGSSPYQTPGIMTMIPHLELGLGTYYPDGGMHR	225
	. . : : ** * : : : * * * . * * : * : : : * . * : * : : : . ** :	
AAA278-O22	LAEKITERCEKLGVDIRLNSRVNEIVLNKSRVTGVVNDATLTFARVVANADAQFVYEKL	294
MbP99-3	ISQSLFELAQKVGVEFRFRKNVKKINHSNNKVTGVTEKGTHTDADIVLCNMDVFPYRQL	285
Nd	ISQSLFELAQKVGKFRFRESVNIITTSKNKVTGVETKNGSYLSDLVVSNDIVPTYRNL	285
	: : : : * . : * : * . : * . . * . : * : * * . . . : * : * * . * . : *	
AAA278-O22	LAPTKKVVNIRKKLAKQEPSLAGFSLLLGLKPSDSQPLEHHTILFPENYDLEFESIFTKR	354
MbP99-3	LQDIK----APEKTLKQERSSSALIFYWGIKKS-FPQLDLHNILFSENYKAEFEAIFNNK	340
Nd	MKDVP----APEKTLKQERSSSALIFYWGIDRE-FPELDLHNILFSENYKTEFEHIFEHK	340
	: : : * . * * : : : * : . . * : * . * * * * : * . * * * * : :	
AAA278-O22	TPVEKPTIYICAPRD-PLMVKDKDHESWFVLVNA PRHSTSGNGFDWSNQDFNRQYANSII	413
MbP99-3	SLYEDPTVYINITSKQSPQDAPKGCENWFVMINTPGD----YGQNWENLVIKA--KKNIL	394
Nd	TLAQDPTVYINITSKESSNDAPAGHENWFVMINAPGD----YGQDWEQLVEES--KKQII	394
	: : * : * * . . * . * * : * : * . * : * . : : * :	
AAA278-O22	NQI-ETAGISIRERLEVLEIRTPLDLQESVNAPGGSIIYGTSSNGARSAFARAKN-RSPIK	471
MbP99-3	SKIKRCLNIDVEELIDVEYVLTPOGIEKNTSSYRGALYGAASNKFAAFLRHPNFNKTIG	454
Nd	AKIKKCLHVDISKHITTEYILTPOGIEKNTSSYRGALYGAASNKFAAFLRHPNFNGKIK	454
	: * . : : : : . : * * : : : : : * : * : * : * . : * * * * . *	
AAA278-O22	GLYLVGGSAAHPGGGLPLVGLSAEMVAKAILE---	502
MbP99-3	NLYHVGGSVHPGGGIPLCLLSAKITADLIPNTNA	488
Nd	NLYHVGGSVHPGGGIPLCLLSAQITADLIQKEQ-	487
	. ** * * * . * * * * : * * * * : . * . * :	