

伊日贵, 祁得林, 朱世海, 杨婷婷, 刘丹, 高强, 夏明哲, 王维, 聂苗苗, 贾军梅, 张存芳 (2026) 基于Cyt *b*序列的高原鳅属鱼类物种鉴定: 系统发育与机器学习方法比较. 生物多样性, 34, 26044. doi: 10.17520/biods.2026044; cstr: 32101.14.biods.2026044.

附录 1 实测数据样本采集信息 ^①

Appendix 1 Sample collection information for empirical data

水系	采样地点	样本编号
Basin	Sampling location	ID
黄河	龙羊峡	L1-L6、L8-L11、L15-L25、L27、L29、L30、L32、L34-L36、 L38-L41、L44、L46、L49-L60、L62、L64-L91、L93-L106
黄河	泽曲河	H1-H90
湟水河	宝库河黑泉水库	M129、M130、M140-M143、M145-M157、M159-M186、 M188、M189、M191-M208、M210-M227、M229-M243、M245- M251、M253-M269、M272、M273
湟水河	北川河西宁段	Y1-Y13
湟水河	宝库河大通县桥头镇	D1-D78

附录 2 样本信息表

Appendix 2 Table of sample information

数据类型	物种名	登录号	单倍型编号	样本数
Data Type	Species Name	GenBank Accession Numbers	Haplotype ID	Sample Size
实测数据	拟硬刺高原鳅 <i>T. pseudoscleroptera</i> (SP1)	PX870651-PX870653、PX870655、	H1-H3、H5、H8、	256
		PX870658、PX870660、PX870663-	H10、H13-H15、	
		PX870665、PX870680-PX870691、	H30-H41、H45、	
		PX870695、PX870696、PX870699、	H46、H49、H52-	
		PX870702-PX870706、PX870709、	H56、H59、H83、	
		PX870733、PX870735	H85	
	斯氏高原鳅 <i>T. stoliczkae</i> (SP2)	PX870654、PX870659、PX870662、	H4、H9、H12、	21
		PX870668-PX870670、PX870672、	H18-H20、H22、	
		PX870673、PX870675、PX870676、	H23、H25、H26、	
		PX870678、PX870679、PX870700、	H28、H29、H50、	
	粗壮高原鳅 <i>T. robusta</i> (SP3)	PX870701	H51	16
		PX870656、PX870657、PX870661、	H6、H7、H11、	
		PX870666、PX870667、PX870671、	H16、H17、H21、	
		PX870692、PX870693	H42、H43	
	黄河高原鳅 <i>T. pappenheimi</i> (SP4)	PX870674、PX870677、PX870697、	H24、H27、H47、	86
		PX870707、PX870708、PX870710-	H57、H58、H60-	
		PX870719、PX870721-PX870724、	H69、H71-H74、	
		PX870727、PX870729、PX870731、	H77、H79、H81、	
	东方高原鳅	PX870732	H82	5
		PX870694、PX870698、PX870734	H44、H48、H84	

① 伊日贵 (2024) 黄河上游主要水域高原鳅复合体的物种鉴定及系统进化研究. 硕士学位论文, 青海大学, 西宁.

伊日贵, 祁得林, 朱世海, 杨婷婷, 刘丹, 高强, 夏明哲, 王维, 聂苗苗, 贾军梅, 张存芳 (2026) 基于Cyt *b*序列的高原鳅属鱼类物种鉴定: 系统发育与机器学习方法比较. 生物多样性, 34, 26044. doi: 10.17520/biods.2026044; cstr: 32101.14.biods.2026044.

数据类型	物种名	登录号	单倍型编号	样本数
Data Type	Species Name	GenBank Accession Numbers	Haplotype ID	Sample Size
下载数据	<i>T. orientalis</i> (SP5) 隆头高原鳅			
	<i>T. alticeps</i> (SP6) 黑体高原鳅	PX870720、PX870728、PX870730	H70、H78、H80	9
	<i>T. obscura</i> (SP7)	PX870725、PX870726	H75、H76	2
	拟硬刺高原鳅	MG697585.1、	GH180、GH181	2
	<i>T. pseudoscleroptera</i> (SP1)	KX373844.1		
		MG697582.1、	GH1-GH60	66
		OP616068.1 -OP616074.1、		
	斯氏高原鳅	MG725404.1-MG725410.1、		
	<i>T. stoliczkae</i> (SP2)	MG725388.1-MG725400.1、		
		MG725380.1-MG725384.1、		
		KX373851.1、		
		MK655046.1-MK655075.1、		
		NC_017890.1、JQ663847.1		
	粗壮高原鳅	OP616095.1、	GH107-GH179	123
		MG697367.1-MG697378.1、		
		MG697384.1、		
	<i>T. robusta</i> (SP3)	MG697386.1-MG697431.1、		
		MG697441.1、		
		MG697443.1-MG697445.1、	GH68-GH80	20
		MG697457.1、MG697525.1-		
		MG697581.1、KX373850.1		
		MG697475.1、MG697476.1、	GH83-GH106	27
	黄河高原鳅	MG697458.1、		
	<i>T. pappenheimi</i> (SP4)	MG697446.1-MG697456.1、		
		MG697385.1、		
		MG697379.1 -MG697383.1、		
	东方高原鳅	MG725414.1、OQ437225.1、		
	<i>T. orientalis</i>	OP616088.1 -OP616094.1、		

伊日贵, 祁得林, 朱世海, 杨婷婷, 刘丹, 高强, 夏明哲, 王维, 聂苗苗, 贾军梅, 张存芳 (2026) 基于Cyt *b*序列的高原鳅属鱼类物种鉴定: 系统发育与机器学习方法比较. 生物多样性, 34, 26044. doi: 10.17520/biods.2026044; cstr: 32101.14.biods.2026044.

数据类型	物种名	登录号	单倍型编号	样本数
Data Type	Species Name	GenBank Accession Numbers	Haplotype ID	Sample Size
	(SP5)	OP616096.1、KX373846.1、 KJ631323.1、KT213599.1、 NC_030505.1、MK655271.1- MK655283.1		
	隆头高原鳅 <i>T. alticeps</i>	OP616075.1-OP616079.1、 KX373837.1、KT213585.1	GH61-GH67	7
	(SP6)			
	黑体高原鳅 <i>T. obscura</i>	MG238305.1、MG238304.1	GH81、GH82	2
	(SP7)			

附录 3 实测数据集单倍型及样本编号
Appendix 3 Haplotypes & sample IDs for empirical dataset

单倍型编号	样本数	样本编号
Haplotype ID	Sample size	Sample ID
H1	25	D1 D5 D35 D48 D59 D73 M156 M175 M185 M189 M191 M192 M195 M198 M207 M216 M221 M223 M226 M236 M238 M241 M260 M268 Y8 D2 D4 D8 D12 D14 D21 D23 D24 D28 D29 D30 D31 D37 D38 D45 D46 D47 D52 D56 D57 D67 D70 D72 D77 L54 M130 M147 M149 M150 M151 M153 M155 M157 M159 M160 M161 M164 M165 M167 M168 M171 M172 M176 M179 M180 M181 M183 M184 M193 M200 M201 M202 M204 M205 M211 M212 M213 M229 M232 M233 M234 M239 M240 M242 M248 M249 M253 M255 M261 M262 M264 M265 M267 M269 M273 Y4 Y11 Y12 Y13
H2	79	
H3	2	D3 M166
H4	1	D6
H5	9	D7 D16 M182 M188 M197 M199 M245 M263 Y5
H6	8	D9 D19 D32 D40 D41 D60 D65 D66
H7	2	D10 D53
H8	1	D11
H9	1	D13
H10	19	D15 D22 D43 D49 D50 D55 D63 L98 L102 M129 M148 M162 M174 M194 M203 M218 M246 M257 Y3
H11	1	D17
H12	1	D18
H13	9	D20 D51 M178 M196 M214 M222 M230 M258 Y6
H14	5	D25 M206 M210 M217 M259
H15	1	D26
H16	1	D27
H17	1	D33

伊日贵, 祁得林, 朱世海, 杨婷婷, 刘丹, 高强, 夏明哲, 王维, 聂苗苗, 贾军梅, 张存芳 (2026) 基于Cyt *b*序列的高原鳅属鱼类物种鉴定: 系统发育与机器学习方法比较. 生物多样性, 34, 26044. doi: 10.17520/biods.2026044; cstr: 32101.14.biods.2026044.

单倍型编号	样本数	样本编号
Haplotype ID	Sample size	Sample ID
H18	1	D34
H19	1	D36
H20	5	D39 D42 D64 D75 M256
H21	1	D44
H22	1	D54
H23	2	D58 D69
H24	3	D61 M220 M225
H25	3	D62 D76 M186
H26	1	D68
H27	1	D71
H28	1	D74
H29	1	D78
H30	18	L1 L2 L5 L6 L25 L27 L32 L35 L38 L64 L74 L75 L88 L89 L90 L91 L101 L106
H31	26	L3 L4 L8 L15 L18 L19 L23 L29 L39 L41 L44 L49 L52 L57 L60 L67 L68 L72 L73 L76 L77 L94 L96 L97 L99 L105
H32	9	L9 L16 L17 L24 L34 L46 L51 L80 L103
H33	15	L10 L20 L22 L36 L40 L50 L53 L55 L56 L65 L69 L78 L93 L95 L100
H34	14	L11 L21 L30 L58 L59 L62 L66 L70 L79 L82 L85 L86 L87 L104
H35	1	L71
H36	1	L81
H37	1	L83
H38	1	L84
H39	1	M140
H40	1	M141
H41	1	M142
H42	1	M143
H43	1	M145
H44	1	M146
H45	1	M152
H46	4	M154 M173 M219 M224
H47	1	M163
H48	3	M169 M215 M227
H49	1	M170
H50	1	Y10
H51	1	M208
H52	2	M231 M237
H53	1	M235
H54	2	M243 M272
H55	1	M247
H56	1	M250

伊日贵, 祁得林, 朱世海, 杨婷婷, 刘丹, 高强, 夏明哲, 王维, 聂苗苗, 贾军梅, 张存芳 (2026) 基于Cyt *b*序列的高原鳅属鱼类物种鉴定: 系统发育与机器学习方法比较. 生物多样性, 34, 26044. doi: 10.17520/biods.2026044; cstr: 32101.14.biods.2026044.

单倍型编号	样本数	样本编号
Haplotype ID	Sample size	Sample ID
H57	1	M251
H58	1	M254
H59	2	M266 Y2
H60	5	H1 H27 H29 H70 H88
H61	32	H2 H3 H5 H6 H7 H12 H15 H16 H20 H22 H24 H31 H36 H38 H43 H49 H51 H53 H60 H62 H65 H68 H71 H72 H74 H76 H81 H83 H85 H86 H87 H90
H62	2	H4 H30
H63	18	H8 H10 H13 H17 H18 H23 H26 H28 H37 H44 H50 H55 H56 H58 H61 H66 H80 H89
H64	1	H9
H65	4	H11 H21 H63 H82
H66	1	H14
H67	1	H19
H68	1	H25
H69	4	H32 H40 H41 H69
H70	7	H33 H47 H52 H54 H57 H75 H78
H71	1	H34
H72	2	H35 H67
H73	1	H39
H74	1	H42
H75	1	H45
H76	1	H46
H77	2	H48 H73
H78	1	H59
H79	1	H64
H80	1	H77
H81	1	H79
H82	1	H84
H83	1	Y1
H84	1	Y7
H85	1	Y9

附录 4 下载数据集单倍型及 GenBank 登录号
Appendix 4 Haplotypes & GenBank accession nos in downloaded dataset

单倍型编号	样本数	GenBank 登录号
Haplotype ID	Sample size	GenBank accession Nos.
GH1	1	MG697582.1
GH2	1	OP616074.1
GH3	1	OP616073.1
GH4	1	OP616072.1
GH5	2	OP616071.1 MG725382.1

伊日贵, 祁得林, 朱世海, 杨婷婷, 刘丹, 高强, 夏明哲, 王维, 聂苗苗, 贾军梅, 张存芳 (2026) 基于Cyt *b*序列的高原鳅属鱼类物种鉴定: 系统发育与机器学习方法比较. 生物多样性, 34, 26044. doi: 10.17520/biods.2026044; cstr: 32101.14.biods.2026044.

单倍型编号	样本数	GenBank 登录号
Haplotype ID	Sample size	GenBank accession Nos.
GH6	1	OP616070.1
GH7	1	OP616069.1
GH8	1	OP616068.1
GH9	1	MG725410.1
GH10	3	MG725409.1 MG725388.1 MK655046.1
GH11	1	MG725408.1
GH12	1	MG725407.1
GH13	1	MG725406.1
GH14	1	MG725405.1
GH15	1	MG725404.1
GH16	1	MG725400.1
GH17	1	MG725399.1
GH18	1	MG725398.1
GH19	1	MG725397.1
GH20	1	MG725396.1
GH21	1	MG725395.1
GH22	1	MG725394.1
GH23	1	MG725393.1
GH24	2	MG725392.1 MK655070.1
GH25	2	MG725391.1 MK655047.1
GH26	1	MG725390.1
GH27	1	MG725389.1
GH28	1	MG725384.1
GH29	1	MG725383.1
GH30	1	MG725381.1
GH31	1	MG725380.1
GH32	1	KX373851.1
GH33	1	MK655075.1
GH34	1	MK655074.1
GH35	1	MK655073.1
GH36	1	MK655072.1
GH37	1	MK655071.1
GH38	1	MK655069.1
GH39	1	MK655068.1
GH40	1	MK655067.1
GH41	1	MK655066.1
GH42	1	MK655065.1
GH43	1	MK655064.1
GH44	1	MK655063.1
GH45	1	MK655062.1
GH46	1	MK655061.1

伊日贵, 祁得林, 朱世海, 杨婷婷, 刘丹, 高强, 夏明哲, 王维, 聂苗苗, 贾军梅, 张存芳 (2026) 基于Cyt *b*序列的高原鳅属鱼类物种鉴定: 系统发育与机器学习方法比较. 生物多样性, 34, 26044. doi: 10.17520/biods.2026044; cstr: 32101.14.biods.2026044.

单倍型编号	样本数	GenBank 登录号
Haplotype ID	Sample size	GenBank accession Nos.
GH47	1	MK655060.1
GH48	1	MK655059.1
GH49	1	MK655058.1
GH50	1	MK655057.1
GH51	1	MK655056.1
GH52	1	MK655055.1
GH53	1	MK655054.1
GH54	1	MK655053.1
GH55	1	MK655052.1
GH56	1	MK655051.1
GH57	1	MK655050.1
GH58	1	MK655049.1
GH59	1	MK655048.1
GH60	2	NC_017890.1 JQ663847.1
GH61	1	KT213585.1
GH62	1	KX373837.1
GH63	1	OP616075.1
GH64	1	OP616076.1
GH65	1	OP616077.1
GH66	1	OP616078.1
GH67	1	OP616079.1
GH68	1	MG697476.1
GH69	1	MG697475.1
GH70	1	MG697458.1
GH71	2	MG697456.1 MG697455.1
GH72	1	MG697454.1
GH73	1	MG697453.1
GH74	3	MG697452.1 MG697449.1 MG697448.1
GH75	1	MG697451.1
GH76	1	MG697450.1
GH77	2	MG697447.1 MG697446.1
GH78	2	MG697385.1 MG697380.1
GH79	1	MG697383.1
GH80	3	MG697382.1 MG697381.1 MG697379.1
GH81	1	MG238305.1
GH82	1	MG238304.1
GH83	1	OP616096.1
GH84	1	OP616094.1
GH85	1	OP616093.1
GH86	1	OP616092.1

伊日贵, 祁得林, 朱世海, 杨婷婷, 刘丹, 高强, 夏明哲, 王维, 聂苗苗, 贾军梅, 张存芳 (2026) 基于Cyt *b*序列的高原鳅属鱼类物种鉴定: 系统发育与机器学习方法比较. 生物多样性, 34, 26044. doi: 10.17520/biods.2026044; cstr: 32101.14.biods.2026044.

单倍型编号	样本数	GenBank 登录号
Haplotype ID	Sample size	GenBank accession Nos.
GH87	1	OP616091.1
GH88	1	OP616090.1
GH89	1	OP616089.1
GH90	1	OP616088.1
GH91	2	OQ437225.1 MK655277.1
GH92	1	MG725414.1
GH93	1	KX373846.1
GH94	1	MK655283.1
GH95	1	MK655282.1
GH96	1	MK655281.1
GH97	1	MK655280.1
GH98	1	MK655279.1
GH99	1	MK655278.1
GH100	1	MK655276.1
GH101	1	MK655275.1
GH102	1	MK655274.1
GH103	1	MK655273.1
GH104	3	MK655272.1 NC_030505.1 KJ631323.1
GH105	1	MK655271.1
GH106	1	KT213599.1
GH107	1	OP616095.1
		MG697581.1 MG697575.1 MG697571.1 MG697570.1 MG697568.1 MG697565.1
GH108	13	MG697553.1 MG697528.1 MG697526.1 MG697418.1 MG697416.1 MG697411.1
		MG697394.1
GH109	2	MG697580.1 MG697576.1
GH110	1	MG697579.1
GH111	3	MG697578.1 MG697534.1 MG697533.1
GH112	1	MG697577.1
		MG697574.1 MG697569.1 MG697564.1 MG697562.1 MG697558.1 MG697557.1
GH113	8	MG697554.1 MG697541.1
GH114	5	MG697573.1 MG697572.1 MG697567.1 MG697566.1 MG697560.1
GH115	2	MG697563.1 MG697552.1
GH116	1	MG697561.1
GH117	1	MG697559.1
GH118	1	MG697556.1
GH119	1	MG697555.1
GH120	1	MG697551.1
GH121	1	MG697550.1
GH122	1	MG697549.1
GH123	1	MG697548.1

伊日贵, 祁得林, 朱世海, 杨婷婷, 刘丹, 高强, 夏明哲, 王维, 聂苗苗, 贾军梅, 张存芳 (2026) 基于Cyt *b*序列的高原鳅属鱼类物种鉴定: 系统发育与机器学习方法比较. 生物多样性, 34, 26044. doi: 10.17520/biods.2026044; cstr: 32101.14.biods.2026044.

单倍型编号	样本数	GenBank 登录号
Haplotype ID	Sample size	GenBank accession Nos.
GH124	1	MG697547.1
GH125	1	MG697546.1
GH126	1	MG697545.1
GH127	1	MG697544.1
GH128	1	MG697543.1
GH129	5	MG697542.1 MG697531.1 MG697529.1 MG697527.1 MG697525.1
GH130	2	MG697540.1 MG697538.1
GH131	1	MG697539.1
GH132	1	MG697537.1
GH133	1	MG697536.1
GH134	2	MG697535.1 MG697530.1
GH135	1	MG697532.1
GH136	2	MG697457.1 MG697431.1
GH137	2	MG697445.1 MG697443.1
GH138	1	MG697444.1
GH139	1	MG697441.1
GH140	1	MG697430.1
GH141	1	MG697429.1
GH142	1	MG697428.1
GH143	1	MG697427.1
GH144	2	MG697426.1 MG697415.1
GH145	2	MG697425.1 MG697422.1
GH146	1	MG697424.1
GH147	1	MG697423.1
GH148	1	MG697421.1
GH149	1	MG697420.1
GH150	1	MG697419.1
GH151	3	MG697417.1 MG697392.1 MG697387.1
GH152	1	MG697414.1
GH153	1	MG697413.1
GH154	2	MG697412.1 MG697395.1
GH155	1	MG697410.1
GH156	1	MG697409.1
GH157	1	MG697408.1
GH158	1	MG697407.1
GH159	1	MG697406.1
GH160	1	MG697405.1
GH161	1	MG697404.1
GH162	7	MG697403.1 MG697401.1 MG697399.1 MG697393.1 MG697388.1 MG697378.1 MG697371.1

伊日贵, 祁得林, 朱世海, 杨婷婷, 刘丹, 高强, 夏明哲, 王维, 聂苗苗, 贾军梅, 张存芳 (2026) 基于Cyt *b*序列的高原鳅属鱼类物种鉴定: 系统发育与机器学习方法比较. 生物多样性, 34, 26044. doi: 10.17520/biods.2026044; cstr: 32101.14.biods.2026044.

单倍型编号	样本数	GenBank 登录号
Haplotype ID	Sample size	GenBank accession Nos.
GH163	2	MG697402.1 MG697400.1
GH164	1	MG697398.1
GH165	1	MG697397.1
GH166	1	MG697396.1
GH167	1	MG697391.1
GH168	1	MG697390.1
GH169	1	MG697389.1
GH170	1	MG697386.1
GH171	1	MG697384.1
GH172	3	MG697377.1 MG697374.1 MG697368.1
GH173	1	MG697376.1
GH174	2	MG697375.1 MG697372.1
GH175	1	MG697373.1
GH176	1	MG697370.1
GH177	1	MG697369.1
GH178	1	MG697367.1
GH179	1	KX373850.1
GH180	1	MG697585.1
GH181	1	KX373844.1

附录 5 不同 k 值下 WEKA 分类器的的鉴定效能对比
Appendix 5 Comparison of identification efficacy of WEKA classifiers under different k values

SMO				J48				JRip				Naïve Bayes				
	CCI(%)	MAE	RMSE	运行时间 Running time	CCI(%)	MAE	RMSE	运行时间 Running time	CCI(%)	MAE	RMSE	运行时间 Running time	CCI(%)	MAE	RMSE	运行时间 Running time
k=4	100	0.2041	0.3012	0.34	97.7444	0.0064	0.0803	0.07	98.4962	0.0046	0.0657	0.29	98.1203	0.0048	0.0662	0.01
k=5	100	0.2041	0.3012	0.47	99.6241	0.0011	0.0328	0.14	98.1203	0.0063	0.0733	0.55	96.6165	0.0097	0.0941	0.01
k=6	100	0.2041	0.3011	0.69	98.8722	0.0032	0.0568	0.52	98.4962	0.0048	0.0662	1.14	98.4962	0.0043	0.0655	0.01
k=7	100	0.2041	0.3012	1.35	99.6241	0.0011	0.0328	1.12	98.8722	0.0036	0.0566	3.41	98.8722	0.0032	0.0559	0.01
k=8	100	0.2041	0.3012	3.49	99.6241	0.0011	0.0328	3.08	98.8722	0.0034	0.0565	12.52	97.3684	0.0102	0.0796	0.01

CCI: 正确分类序列占比(%); MAE: 平均绝对误差; RMSE: 均方根误差。
CCI, Correctly Classified Instances(%); MAE, Mean Absolute Error; RMSE, Root Mean Squared Error.