[23/09/2013]

Re-submitted to *Journal of Molecular Evolution*

Research article

Tandem duplications in the C-terminal domain of the mesotocin receptor exclusively identified among East Eurasian thrushes

Hideaki Abe[1, 3], Isao Nishiumi[2] and Miho Inoue-Murayama[1, *]

[1]*Wildlife Research Center, Kyoto University, Kyoto, 606-8203, Japan*

[2]*Department of Zoology, National Museum of Nature and Science, Tsukuba, 305-0005, Japan*

[3] Current address:
Department of Anatomy, University of Otago, Dunedin, 9054, New Zealand

[*]Correspondence to:
Miho Inoue-Murayama
Wildlife Research Center of Kyoto University
2-24 Tanaka-Sekiden-cho, Sakyo-ku, Kyoto, 606-8203, Japan
Telephone: +81-75-751-4375 Fax:+81-75-771-4394
e-mail: mmurayama@wrc.kyoto-u.ac.jp

Abstract

Mesotocin is a neurohypophyseal hormone found in some non-mammalian vertebrates, including birds, reptiles and amphibians. In this study, we identified and characterized 18-amino acid duplications in the C-terminal domain of the mesotocin receptor (MTR), specifically found in *Turdus* thrushes (Aves: Passeriforms: Turdidae). These duplicated elements are located in the distal part of the C-terminal tails of MTR and consist of amino acids that are highly conserved among major vertebrates. Intraspecific polymorphisms in a variable number of tandem duplications are commonly found in East Eurasian *Turdus*, but not in any other genus of Turdidae. Moreover, the genus *Turdus* can be further classified into 2 groups according to the presence or absence of a 3-amino acid deletion just adjacent to the putative palmitoylation site in the cytoplasmic C-terminal tail. The phylogeny presented here strongly supports the conspecific group of 4 East Eurasian thrushes (*Turdus pallidus*, *T. chrysolaus*, *T. obscurus*, and *T. celaenops*). Our findings, therefore, provide a new synapomorphy that can be used for phylogenetic assumptions and shed a light on the history of diversification within Eurasian *Turdus* clades.

Keywords: mesotocin receptor — C-terminal domain — tandem duplication — polymorphism — insertions and deletions — *Turdus*

54 Introduction

55

56 Mesotocin (MT) is the non-mammalian homolog of mammalian oxytocin, and this neurohypophysial
57 hormone exerts biological effects by binding to G protein-coupled receptors (GPCRs) on the cell
58 surface. GPCRs are structurally characterized by 7 transmembrane domains, intracellular and
59 extracellular loops, and amino- and C-terminal segments (Gimpl and Fahrenholz 2001; Lolait et al.
60 1992). We focused on length variations in the C-terminal domain of peptide hormone receptor as a
61 key factor in influencing signal transduction pathways. Hoare et al. (1999) experimentally illustrated
62 that artificial truncation in the C-terminal tail of the human oxytocin receptor has critical effects in
63 receptor-G protein interactions. Furthermore, our recent study demonstrated that the arginine
64 vasotocin receptor has inter- and intraspecific variations in the length of C-terminal domains,
65 suggesting conformational differences in this region (Abe et al. 2012). Thus, we consider it worth
66 investigating whether the mesotocin receptor (MTR) also has conserved sequence motifs and amino
67 acid residues in the cytoplasmic C-tail.
68 In addition to characterization of novel polymorphisms, in this study we also present a hypothesis of
69 phylogenetic relationships among East Eurasian thrushes by taking advantage of synapomorphic
70 feature of insertions and deletions (indels) that are specifically found in *Turdus* clades. A
71 synapomorphic character is one of the most reliable piece of information on phylogenetic

72 reconstitution, as long as parallel changes and/or reversions are unlikely between different taxa. Not
73 only morphological characteristics but also genomic changes can be used as a landmark for
74 identifying a taxonomic clade. Rare genomic changes (RGCs) occasionally provide unambiguous
75 evidences for resolving the puzzle of phylogenetic problems (reviewed in Boore 2006; Rokas and
76 Holland 2000). RGCs can be caused by various molecular events—retroposon integrations, gene
77 order changes, and large-scale indels, etc. Some retroposons known as long and short interspersed
78 repetitive elements (LINEs and SINEs, respectively) have attracted special attention of evolutionary
79 biologists, because copies of these molecules shared at the same locus in 2 or more different taxa are
80 derived from the same element originally inserted into the germline of a common ancestor (Shedlock
81 and Okada 2000). Indeed, the presence or absence of SINEs insertions have been used for
82 phylogenetic analyses in a variety of organisms such as salmons (Kido et al. 1991), cetaceans
83 (Nikaido et al. 1999), and avian species (St. John et al. 2005; Watanabe et al. 2006). Moreover, some
84 amino acid indels and substitutions that had occurred in functional proteins proved to be
85 synapomorphic characters for avian clades (Ericson et al. 2000; Groth and Barrowclough 1999;
86 Stapel et al. 1984).
87 The genus *Turdus* is the largest group of Turdidae and perhaps the most successful of any passerine
88 genus in the world in terms of natural geographical penetration and degree of speciation. *Turdus* is
89 the only thrush genus that is widely distributed throughout Eurasia, South and North America, and

90 Africa. Of its nearly 70 species, 8 are confined to Africa, 25 are essentially Asian, and 34 are
91 restricted to the America (Collar 2005). A previous study on phylogenetic relationships among true
92 thrushes indicated that *Turdus* radiations occurred near the end of the Miocene (7–8 million years
93 ago) and rapidly diversified into each continental lineage (Klicka et al. 2005). Although the
94 subsequent statistical analysis based on mitochondrial genes indicated a monophyly of each
95 continental clade, relatively weak nodal support was found within the Eurasian *Turdus* clade
96 (Voelker et al. 2007). Such a rapid radiation and speciation, which are general characteristics of
97 avian evolution, always hamper phylogenetic approaches using stochastic methods (see Poe and
98 Chubb 2004). Another striking feature of the genus *Turdus* is its apparent plumage
99 homogeneity—similar plumage patterns appear repeatedly across its wide range. This makes it more
100 complicated to define species boundaries and unravel the evolutionary history. Therefore, it is
101 important to gather information on RGCs, because one can obtain invaluable data sets for cladistic
102 assumptions if synapomorphic characters could be found elsewhere within problematic taxa. This
103 approach is expected to provide a better resolution of phylogenetic relationships and interesting
104 insights into speciation processes and timing at multiple taxonomic levels.
105
106
107 Material and methods

Avian specimens and DNA extraction

A wide range of avian species were used in this study to explore sequence and structural features of the C-terminal tails of the MTR. In addition, more than 200 *Turdus* specimens were particularly used to further investigate intraspecific polymorphisms in this domain of the receptor. These specimens were divided into 3 categories, depending on their sources: (1) live resources derived from zoos and aviaries, (2) carcasses originated either from wildlife conservation acts or from bird-strike events in Japanese airports, and (3) specimens obtained from avian research projects conducted by the National Museum of Nature and Science. Supplementary S-Table 1 summarizes the avian samples used for sequence alignment and their GenBank IDs. When we collected specimens from live animals, we strictly observed the domestic laws for animal rights and protection in Japan. DNA was extracted from various materials such as feathers, blood, and tissues (e.g., muscle or liver) using either of the 3 methods: DNeasy Blood and Tissue Kit (Qiagen, Tokyo, Japan), Puregene DNA Purification Kit (Gentra Systems, Minneapolis, MN, USA), or standard phenol-chloroform extraction (Sambrook et al. 1989).

Polymerase chain reaction amplification

126

127 To amplify the C-terminal region of the MTR, primers were designed on the basis of the nucleotide

128 sequences of the zebra finch deposited in GenBank (Accession No. XM_00218826). The following

129 oligo primers were designed for amplification of avian *MTR*: MTf2 (5′–

130 GCCTCCCCCTTCATCATCG–3′), MTr3 (5′–AAGCTCCTGTGGCTCGTG–3′), MTr4 (5′–

131 CGGCAGCTGAGCACGAAG–3′), and MTr7 (5′–TCGCGGGCGGCTGCACGAA–3′). We mainly

132 used the MTf2/MTr3 primer pair, but MTf2/MTr4 was employed for amplification of Phasianidae

133 (Galliformes) and Anatidae (Anseriformes). MTf2/MTr7 was used in the amplifications of *Turdus*

134 thrushes, because the MTr3 annealing site overlapped with lineage-specific tandem duplicated

135 elements. In each case, polymerase chain reaction (PCR) was carried out in a PCR cocktail

136 containing approximately 10–50 ng of genomic DNA, 0.25 μM each primer, 0.2 mM dNTPs, 5 ×

137 tuning buffer, and 0.375 U of G-Taq DNA polymerase (Hokkaido System Science, Hokkaido, Japan).

138 PCR conditions comprised an initial incubation at 95°C for 2 min, followed by 40 cycles of 95°C for

139 15 s and 60°C for 30 s.

140

141 *Sequencing and multiple alignments*

142

143 The nucleotide sequences of the C-terminal domain of MTR were determined in a total of 44 avian

144 species belonging to 31 families (17 orders). Sequencing was performed on an ABI 3130xl
145 automated sequencer (Applied Biosystems, Foster City, CA, USA) using BigDye Terminator Cycle
146 Sequencing reagents v3.1, according to the manufacturer's instructions (Applied Biosystems).
147 Nucleotide sequences were translated into amino acids and then aligned by using CLUSTAL W
148 (Thompson et al. 1994) implemented in BioEdit ver. 7.9.9.0 (Hall 1999). Each band of the amplified
149 alleles was purified from the gel and subjected to DNA sequencing. Information on amino acid
150 sequences of the C-terminal tail of vertebrate oxytocin and MT receptors were collected either from
151 previous papers (Akhundova et al. 1996; Hoare et al. 1999) or from GenBank sequences.
152
153 *Genotyping and sex determination*
154
155 Multiple sequence alignment revealed the existence of tandem duplicated elements in the distal
156 C-terminal tails of MTR specifically in *Turdus* thrushes. Thus, the pattern of length polymorphisms
157 was further investigated by genotype analysis using capillary electrophoresis. MTf2 was
158 fluorescently labeled with 6-carboxyfluorescein (6-FAM), and the HD500 Rox size standard
159 (Applied Biosystems) was used as an internal standard. PCR products were electrophoresed on an
160 ABI 3130xl sequencer, followed by peak detection with the Genemapper software (Applied
161 Biosystems). When we could not judge the gender according to the external features of the cloaca,

sexes were determined using the DNA-based identification method described by Griffiths et al. (1998). The exact fragment size of partial CHD-W and CHD-Z on automated capillary electrophoresis was taken from the empirical data of Lee et al. (2010). We evaluated the extent of statistical heterogeneity in allelic distributions between sampling locations (west Honsyu [n = 39] vs. Kyusyu [n = 14]) and sexes (male [n = 55] vs. female [n = 29]) of the pale thrush (*Turdus pallidus*) using Fisher's exact test.

Results

Structural characteristics in the C-terminal domain of the avian mesotocin receptor

In this study we determined the nucleotide sequences of the C-terminal domain of MTR in 44 avian species belonging to 17 orders. These sequences were deposited in the GenBank database (Accession Nos. AB634778–AB634824, and AB743549). We did not find any stop codon or frame-shift mutation in the sequence alignment; hence, we excluded the possibility that these sequences could be derived from MTR pseudogenes. The schematic secondary structure of the target region is shown in Figure 1. Multiple alignments showed that the distal part of TM VII (position

318–335) and its juxtamembrane region (position 336–343) are completely conserved among all avian species (Fig. 2). The conserved NPxxY motif, which is characteristic of rhodopsin GPCRs (Fritze et al. 2003), could be identified in TM VII (position 328–332). A double cysteine motif in the C-terminus (positions 349, 350) was conserved in all of the avian and mammalian species; these residues have been suggested by the CSS-Palm software (Ren et al. 2008) to be putative sites for palmitoylation. Passerine birds could be distinguished from other avian species because of the common deletion of 2 amino acids in the middle of the C-terminal domain (positions 364, 365). Furthermore, we found 54-bp duplicated elements (18 amino acids; KSNSSSFVLSCRSPSHRS) in *Turdus* thrushes. There were 4 types of alleles with the tandem repeat number varying from 1 to 4 [*NR* (wild-type), *RT2*, *RT3*, and *RT4*]. Four alleles were different by exactly 60-bp intervals. Global alignment with other vertebrates uncovered that the duplicated region was highly conserved not only among avian species but also among the mammalian species (Fig. 3). In particular, the KSXSSXFVLS (X: arbitrary amino acid) motif was completely conserved in all vertebrates except teleost fish. Although the functional significance of repeated motifs remains speculative, it should be mentioned that tandem duplicated motifs contain consensus sites for phosphorylation by the multifunctional protein kinase II (RxxS; Kemp and Pearson 1990).

Distribution of tandem duplicated alleles within Turdus thrushes

We detected intraspecific variations in the number of tandem duplications among *Turdus* thrushes, whereas the White's thrush (*Zoothera dauma*) was monomorphic with the wild-type (i.e., non-duplicated) allele (Table 1). The other following avian species tested for length polymorphisms did not show any variation in this locus (zebra finch [*Taeniopygia guttata*; *n* = 32], Bengalese finch [*Lonchura striata var. domestica*; *n* = 15], red-throated parrot-finch [*Erythrura psittacea*; *n* = 8], canary [*Serinus canaria*; *n* = 10], rock eagle-owl [*Bubo bengalensis*; *n* =8], and mallard [*Anas platyrhynchos*; *n* = 8]). These data indicate that duplication events occurred in the common ancestor of *Turdus* thrushes. There was little difference in the allelic frequency among Eurasian *Turdus* species. *RT2* was the predominant allele with a frequency ranging from 0.62 to 0.74; the frequency of the other alleles was similar among species.

Phylogenetic assumption based on rare genomic changes

In addition to tandem duplications in the C-terminal tails, we found a 3-amino acid deletion (Ser–Thr–Arg) at residues 351 to 353, just adjacent to a putative palmitoylation site (see Figs. 1 and 2). The non-duplicated and duplicated alleles with this deletion were expressed as *NR–*, *RT2–*, *RT3–*, and *RT4–*, and the frequency of each allele is shown in Table 1. The deletion was conserved in all

individuals belonging to 4 *Turdus* species (*T. pallidus*, brown-headed thrush [*T. chrysolaus*], eyebrowed thrush [*T. obscurus*], and Izu thrush [*T. celaenops*]), but no deletion was detected in the other *Turdus* (Naumann's thrush [*T. naumanni*] and Japanese thrush [*T.cardis*]) and outgroup genera *Zoothera* (*Z. dauma*). Therefore, the East Eurasian thrushes can be divided into 3 groups according to the presence or absence of these synapomorphic indels. Eurasian *Turdus* Group 1(ETG1) has characteristics of both tandem duplications and deletions, and includes *T. pallidus* and its conspecies *T. chrysolaus*, *T. obscurus*, and *T. celaenops*. Eurasian *Turdus* Group 2 (ETG2) consists of the other major *Turdus* thrushes (*T. naumanni* and *T.cardis*) whose only tandem duplications are detected in the C-terminal domain. The thrush group who has neither tandem duplications nor the 3-amino acid deletion is considered to form an outer clade of ETG1 and ETG2 (Fig. 4) and named OT (other species of Turdidae).

In the next step, we tried to identify to which group the other *Turdus* species belong to, according to the presence or absence of these synapomorphic indels. Two samples of the grey-backed thrush (*T. hortulorum*) collected in central Japan had the genotype of *RT2*/*RT2* and *RT2*/*RT3*, respectively, thus belonging to ETG2 (Table 2). The other 2 Eurasian *Turdus* species (island thrush [*T. poliocephalus niveiceps*] and rufous-throated thrush [*T. ruficollis*]) also had duplicated alleles that were characteristics of ETG2 thrushes, whereas the Siberian thrush (*T. sibiricus*) could be clearly distinguished from other *Turdus* species by the absence of a duplicated allele.

234

235

236 Discussion

237

238 We detected intraspecific polymorphisms in a variable number of tandem duplications in all *Turdus*
239 species (with large sample sizes), even though we did not detect a significant difference in the allelic
240 distribution between sexes ($P = 0.383$) nor between sampling locations (Honsyu vs. Kyusyu; $P =$
241 0.257) in *T. pallidus* populations. This simply suggests that tandem duplications in the C-terminal
242 tail would have no impact on the receptor functions, and that the current level of allelic distributions
243 is determined by a random genetic drift under the neutral mutation model of evolution. One of the
244 most plausible explanations for the retention of polymorphism within each *Turdus* species is gene
245 conversion events associated with genetic drift. Recent study suggests that gene conversion can
246 generate new haplotypes by transferring sections of DNA within and across duplicated loci in wild
247 bird populations (Spurgin et al. 2011). Similar mechanism may contribute to generate genetic
248 instability in the C-terminal domain of MTR. Gene conversion can also homogenize the level of
249 variation among populations by increasing or decreasing of the tandem duplication unit. At present,
250 however, we do not have an answer to the following fundamental questions: Why can only the
251 genomes of *Turdus* thrushes accommodate structural changes in the C-terminal domain of MTR?

252 Why are duplicated elements so strictly conserved among vertebrates? To address these questions,
253 we need to conduct site-directed mutagenesis and/or other biochemical experiments targeting
254 C-terminal domain of MTR.
255 The phylogenetic relationship inferred from synapomorphic indels strongly supports the conspecific
256 relationships among *T. pallidus*, *T. chrysolaus*, *T. obscurus*, and *T. celaenops*. This is consistent with
257 molecular phylogenies based on the cytochrome c oxidase subunit I (*COI*) sequences (Saitoh et al.
258 manuscript in preparation) and datasets from 3 mitochondrial genes (Voelker et al. 2007). Moreover,
259 local ornithologists have reported that some of Eurasian *Turdus* belonging to EGT1 are temporally
260 intermingled with each other for wintering in Japan, and that *T. chrysolaus* sometimes can be seen in
261 the same flock with *T. pallidus*. All these findings suggest a close relationship and recent
262 diversification of ETG1 thrushes. In fact, captive and natural hybridization was reported between
263 several combinations of closely related *Turdus* species such as *T. chrysolaus* × *T. pallidus* (male) and
264 *T. naumanni* × *T. ruficollis* (McCarthy 2006, p. 239). Natural hybrid zones between Eurasian *Turdus*
265 species may help to maintain the allele frequency constant among each group of *Turdus* thrushes,
266 even though our data presented here clearly reject the possibility of hybridization between ETG1 and
267 ETG2 thrushes. *T. celaenops* is an exceptional species, because it is considered a sedentary bird that
268 breeds and reproduces on small islands, in contrast to the other migratory thrushes. Therefore, *T.*
269 *celaenops* is considered to have lost the rare alleles (*RT3*– and *RT4*–) recently due to local bottleneck

effects in some island populations.

Not only hybridization with closely related species but also other biological factors influence the magnitude of polymorphisms in MTR. *T. pallidus* is a migratory bird of eastern Asia, and southern and central parts of Japan are their primary wintering grounds (Brazil 1991). According to the first breeding record in the Hiroshima prefecture (located in the western part of Honsyu) in 1991, some populations of *T. pallidus* stay in Japan throughout the year (Ueno et al. 1993). Such a social and behavioral dimorphism would promote diversification and speciation in local populations, resulting in stochastic transitions of genetic variation. In our future research, the levels and patterns of polymorphisms in other neurotransmitter-related genes will be examined between resident and migrating populations of *T. pallidus*.

282 Acknowledgments

283

284 The authors wish to thank the following persons and institutions for providing avian specimens:

285 Kayoko Kameda (Lake Biwa Museum), Kimiyuki Tsuchiya (Applied Biology Co. Ltd.), Hitoshi

286 Suzuki (Hokkaido University), Masatsugu Suzuki and Kumi Haraguchi (Gifu University), Masahiko

287 Sakanashi (Matsubase Store House), Mizuki Mizutani (Fukui Nature Conservation Center), Yusaku

288 Watanabe (Iwate Wildlife Rescue Center), Hideyuki Ito (Kyoto City Zoo), Tomoko Kawakami

289 (Kujukushima Zoological and Botanical Garden), Taku Mizuta (Amami Wildlife Center), and Yuji

290 Miyake (Yambaru Wildlife Center). We also thank Azusa Hayano (Kyoto University) for great

291 assistance with laboratory work. This study was supported financially by the Ministry of Education,

292 Culture, Sports, Science and Technology (MEXT) with a Grant-in-aid for Science Research

293 (#21310150 to MI-M) and Global Center of Excellence (GCOE) Program “Formation of a Strategic

294 Base for Biodiversity and Evolutionary Research: from Genome to Ecosystem”

295

Conflict of interest

The authors declare that they have no conflict of interest.

299 References

300

301 Abe H, Watanabe Y, Inoue-Murayama, M (2012) Genetic variation in the C-terminal domain of
302 arginine vasotocin receptor in avian species. Gene 494:174–180.
303 Akhundova A, Getmanova E, Gorbulev V, Carnazzi E, Eggena P, Fahrenholz F (1996) Cloning and
304 functional characterization of the amphibian mesotocin receptor, a member of the
305 oxytocin/vasopressin receptor superfamily. Eur J Biochem 237:759–767.
306 Boore JL (2006) The use of genome-level characters for phylogenetic reconstitution. Trends Ecol
307 Evol 21:439–446.
308 Brazil MA (1991) The bird of Japan. Christopher Helm, London.
309 Collar NJ (2005) Family TURDIDAE (THRUSHES). In: del Hoyo J, Elliott A, Christie DA (eds)
310 Handbook of the birds of the world. vol. 10: Cuckoo-shrikes to Thrushes. Lynx edicions,
311 Balcelona, pp 514–807.
312 Ericson PGP, Johansson US, Parsons TJ (2000) Major divisions in oscines revealed by insertions in
313 the nuclear gene *c-myc*: a novel gene in avian phylogenetics. Auk 117:1069–1078.
314 Fritze O, Filipek S, Kuksa V, Palczewski K, Hofmann KP, Ernst OP (2003) Role of the conserved
315 NPxxY(x)5,6F motif in the rhodopsin ground state and during activation. Proc Natl Acad
316 Sci USA 100:2290–2295.
317 Gimpl G, Fahrenholz F (2001) The oxytocin receptor system: structure, function, and regulation.
318 Physiol Rev 81:629–683.
319 Griffiths R Double MC Orr K Dawson RJ (1998) A DNA test to sex most birds. Mol Ecol 7:1071–
320 1075.
321 Groth JG, Barrowclough GF (1999) Basal divergences in birds and the phylogenetic utility of the
322 nuclear RAG-1 gene. Mol Phylogenet Evol 12:115–123.
323 Hall TA (1999) Bioedit: a user-friendly biological sequence alignment editor and analysis program
324 for Windows 95/98/NT. Nucl Acids Symp Ser 41:95–98.
325 Hoare S, Copland JA, Strakova Z, Ives K, Jeng YJ, Hellmich MR, Soloff MS (1999) The proximal
326 portion of the COOH terminus of the oxytocin receptor is required for coupling to g(q), but
327 not g(i). Independent mechanisms for elevating intracellular calcium concentrations from
328 intracellular stores. J Biol Chem 274:28682–28689.
329 Kemp BE, Pearson RB (1990) Protein kinase recognition sequence motifs. Trends Biochem Sci
330 15:342–346.
331 Kido Y, Aono M, Yamaki T, Matsumoto K, Murata S, Saneyoshi M, Okada N (1991) Shaping and
332 reshaping of salmonid genomes by amplification of tRNA-derived retroposons during

333 evolution. Proc Natl Acad Sci USA 88:2326–2330.
334 Klicka J, Voelker G, Spellman GM (2005) A molecular phylogenetic analysis of the “true thrushes”
335 (Aves: Turdinae). Mol Phylogenet Evol 34:486–500.
336 Lee JC, Tsai LC, Hwa PY, Chan CL, Huang A, Chin SC, Wang LC, Lin JT, Linacre A, Hsieh HM
337 (2010) A novel strategy for avian species and gender identification using the CHD gene.
338 Mol Cell Probes 24:27–31.
339 Lolait SJ, O'Carroll AM, McBride OW, Konig M, Morel A, Brownstein MJ (1992) Cloning and
340 characterization of a vasopressin V2 receptor and possible link to nephrogenic diabetes
341 insipidus. Nature 357:336–339.
342 McCarthy EM (2006) Handbook of Avian Hybrids of the World. Oxford University Press, Oxford.
343 Nikaido M, Rooney AP, Okada N (1999) Phylogenetic relationships among cetaartiodactyls based on
344 evidence from insertions of SINEs and LINEs: hippopotamuses are the closest extant
345 relatives of whales. Proc Natl Acad Sci USA 96:10261–10266.
346 Poe S, Chubband AL (2004) Birds in a bush: five genes indicate explosive evolution of avian orders.
347 Evolution 58:404–415.
348 Ren J, Wen L, Gao X, Jin C, Xue Y, Yao X (2008) CSS-Palm 2.0: an updated software for
349 palmitoylation sites prediction. Protein Eng Des Sel 21:639–644.
350 Rokas A, Holland PWH (2000) Rare genomic changes as a tool for phylogenetics. Trends Ecol Evol
351 15:454–459.
352 Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning. A laboratory manual. Cold Spring
353 Harbor Press, New York.
354 Shedlock AM, Okada N (2000) SINE insertions: powerful tools for molecular syatematics.
355 BioEssays 22:148–160.
356 Spurgin, LG, van Oosterhout C, Illera JC, Bridgett S, Gharbi K, Emerson BC, Richardson DS (2011)
357 Gene conversion rapidly generates major histocompatibility complex diversity in recent
358 founded bird populations. Mol Ecol 20:5213-5225.
359 Stapel SO, Leunissen JAM, Versteeg M, Wattel J, de Jong WW (1984) Ratites as oldest offshoot of
360 avian stem–evidence from α-crystallin A sequences. Nature 311:257–259.
361 St. John J, Cotter J-P, Quinn TW (2005) A recent chicken repeat 1 retrotransposition confirms the
362 Coscoroba–Cape Barren goose clade. Mol Phylogenet Evol 37:83–90.
363 Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of
364 progressive multiple sequence alignment through sequence weighting, position-specific gap
365 penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680.
366 Ueno Y, Totune T, Yanamoto H, Hibino M, Iida T (1993) First breeding record of the pale ouzel
367 *Turdus pallidus* from Honshu. Jpn J Ornithol 41:17–19.
368 Voelker G, Rohwer S, Bowie RCK, Outlaw DC (2007) Molecular systematics of a speciose,

cosmopolitan songbird genus: Defining the limits of, and relationships among, the *Turdus* thrushes. Mol Phylogenet Evol 42:422–434.

Watanabe M, Nikaido M, Tsuda TT, Inoko H, Mindell DP, Murata K, Okada N (2006) The rise and fall of the CR1 subfamily in the lineage leading to penguins. Gene 365:57–66.

Table 1

Distributions of wildtype and duplicated alleles among *Turdus* and *Zoothera* thrushes

Species/allele*	*2n*	*NR*	*NR−*	*RT2*	*RT2−*	*RT3*	*RT3−*	*RT4*	*RT4−*
T. pallidus	170		25 (0.15)		120 (0.71)		22 (0.13)		3 (0.02)
T. chrysolaus	22		6 (0.27)		15 (0.68)		1 (0.05)		—
T. obscurus	10		1 (0.1)		7 (0.7)		1 (0.1)		1 (0.1)
T. celaenops	86		22 (0.26)		64 (0.74)		—		—
T. cardis	52	9 (0.17)		32 (0.62)		9 (0.17)		2 (0.04)	
T. naumanni	24	3 (0.13)		16 (0.67)		3 (0.03)		2 (0.08)	
Zoothera dauma	18	18 (1)		—		—		—	

374

375 * *NR* (wild-type, no duplication), *RT2* (2 tandem repeats), *RT3* (3 tandem repeats), and *RT4* (4 tandem

376 repeats). A dush (–) biside allele name indicates the presence of the 3-amino acid deletion.

377

Table 2

Genotyping in *Turdus* species with less than 3 specimens

Species	*n*	Locations	Genotype(s)*	Group
T. poliocephalus niveiceps	1	Taiwan	*RT2*/*RT2*	ETG2
T. ruficollis	1	Mongolia	*NR*/*RT2*	ETG2
T. hortulorum	2	central Japan	*RT2*/*RT2*, *RT2*/*RT3*	ETG2
T. sibiricus	2	northern Japan	*NR*/*NR*, *NR*/*NR*	OT

* *NR* (wild-type, no duplication), *RT2* (2 tandem repeats) , and *RT3* (3 tandem repeats)

381 Figure Legends

382

383 **Fig. 1**

384 Schematic structure of the C-terminal tail of the mesotocin receptor (MTR). The double cysteine

385 residues for palmitoylation are highlighted and the amino acid residues highly conserved among

386 vertebrates are shown in capital letters. The position of the 3-amino acid deletion (Ser–Thr–Arg) is

387 boxed, and the tandem duplicated domain in *Turdus* thrushes is highlighted with gray background.

388

389 **Fig. 2**

390 Comparison of the amino acid sequences at the C-terminal tail of mammalian oxytocin receptors

391 (OXTRs), MTRs (amphibian, marsupial, and avian), and fish isotocin receptor. GenBank sequences

392 are derived from rat (rOXTR; NP037003.2), human (hOXTR; NP000907.2), cow (cOXTR;

393 NP776559.1), porcine (pOXTR; NP999192.1), canine (dOXTR; NP001185588.1), western clawed

394 frog (fMTR; XM002936297.1), gray short-tailed opossum (oMTR; XP001375059.1), green anole

395 (aMTR; XP003224939.1), non-passerine Neoaves (npMTR; AB634795—AB634816), zebra finch

396 (zfMTR; XM002188266.1), chicken (cMTR; NP001026740.1), and zebrafish (fISR;

397 XP001341507.1). RT2, RT3, and RT4 are duplicated elements found in *Turdus* thrushes. Highly

398 conserved motifs are shaded, and the symbols ¶ and § represent S/G/N/A/T and S/G, respectively.

399 The sequence of npMTR is truncated because of internal primer design.

400

401

402 **Fig. 3**

403 Alignment of amino acid sequences in the distal part of transmembrane VII and the C-terminal tail

404 of the avian mesotocin receptor. A dot or a dash denotes identity with the consensus sequence or a

405 deletion, respectively. The GenBank sequence of zebra finch (XM002188266) is also used for

406 alignment, and conserved residues in all avian taxa are highlighted. X in the sequence of Pthrush

407 [RT3–] indicates alanine or glycine due to heterozygous nucleotides in this position. A deletion that

408 could be identified only in Eurasian *Turdus* Group 1 (ETG1) is highlighted with gray background.

409 The square brackets indicate the number of repetitive units in the C-terminal tail, and duplicated

410 elements are shown at the bottom of the alignment (each repeat unit is boxed). Vertical bars represent

411 the following categories: P, passerine birds; N, non-passerine Neoaves; G, Galliformes; and A,

412 Anseriformes. Details of avian specimens are shown in S-Table 1.

413

414 **Fig. 4**

415 Phylogenetic relationships of East Eurasian thrushes inferred from synapomorphic characters at the

416 C-terminal domain of the mesotocin receptor. The timings of 3-amino acid deletion and 18-amino

417 acid duplications are indicated as red circle and boxes, respectively. Eurasian *Turdus* thrushes can be

418 divided into 2 groups, Eurasian *Turdus* Groups 1 and 2 (ETG1 and ETG2) according to these

419 synapomorphic characters.

420

421 Supplementary material

422 **Table S1** Details of avian specimens used for sequencing

423

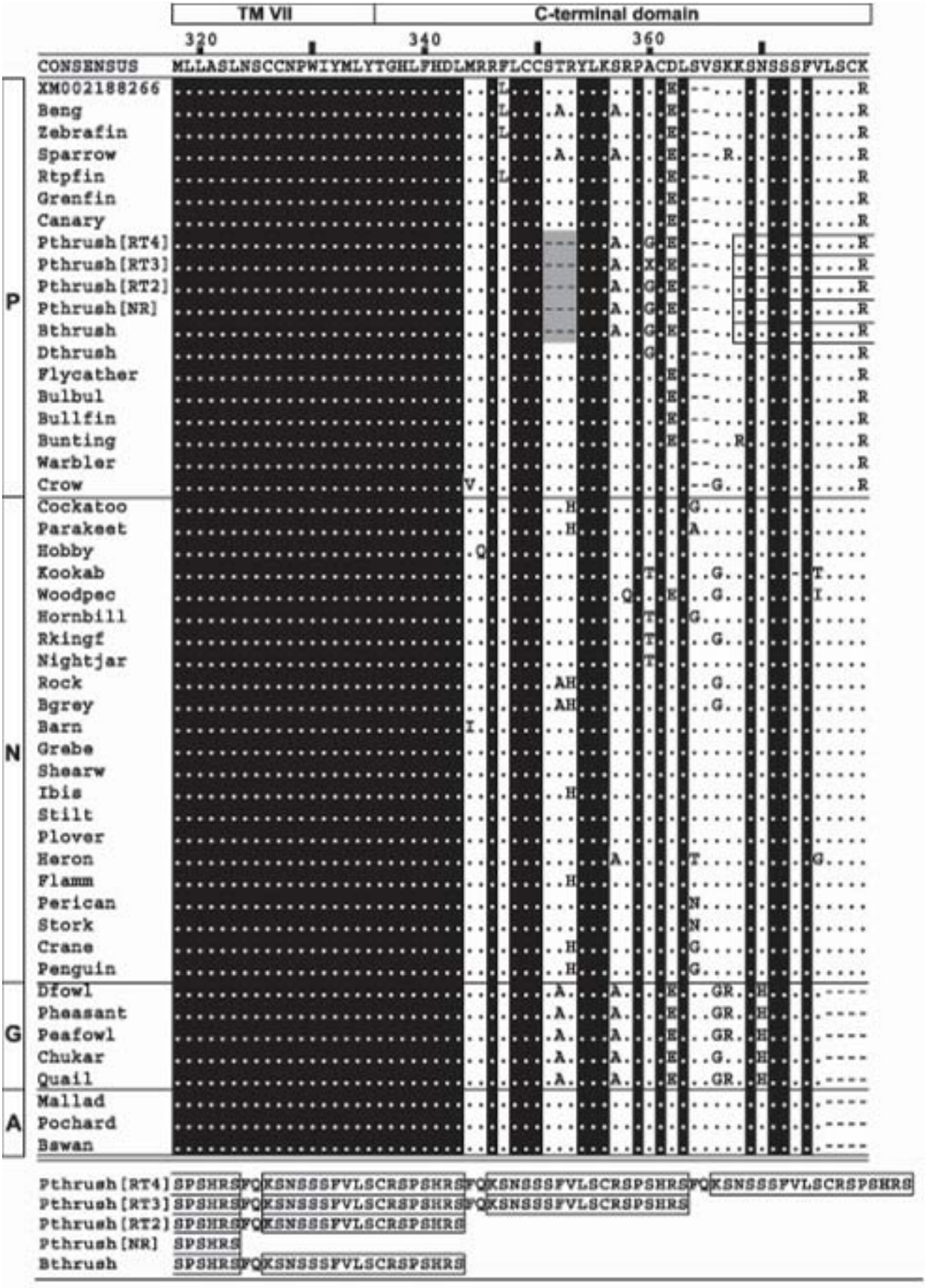
TM VII
C-terminal domain
320
340
360
CONSENSUS MLLASLNSCCNPWIYMLYTGHLFHDLMRRFLCCSTRYLKSRPACDLSVSKKSNSSSFVLSCK
P
XM002188266
Beng
Zebrafin
Sparrow
Rtpfin
Grenfin
Canary
Pthrush[RT4]
Pthrush[RT3]
Pthrush[RT2]
Pthrush[NR]
Bthrush
Dthrush
Flycather
Bulbul
Bullfin
Bunting
Warbler
Crow
N
Cockatoo
Parakeet
Hobby
Kookab
Woodpec
Hornbill
Rkingf
Nightjar
Rock
Bgrey
Barn
Grebe
Shearw
Ibis
Stilt
Plover
Heron
Flamm
Perican
Stork
Crane
Penguin
G
Dfowl
Pheasant
Peafowl
Chukar
Quail
A
Mallad
Pochard
Bswan
Pthrush[RT4] SPSHRSFQKSNSSSFVLSCRSPSHRSFQKSNSSSFVLSCRSPSHRSFQKSNSSSFVLSCRSPSHRS
Pthrush[RT3] SPSHRSFQKSNSSSFVLSCRSPSHRSFQKSNSSSFVLSCRSPSHRS
Pthrush[RT2] SPSHRSFQKSNSSSFVLSCRSPSHRS
Pthrush[NR] SPSHRS
Bthrush SPSHRSFQKSNSSSFVLSCRSPSHRS

Fig. 2

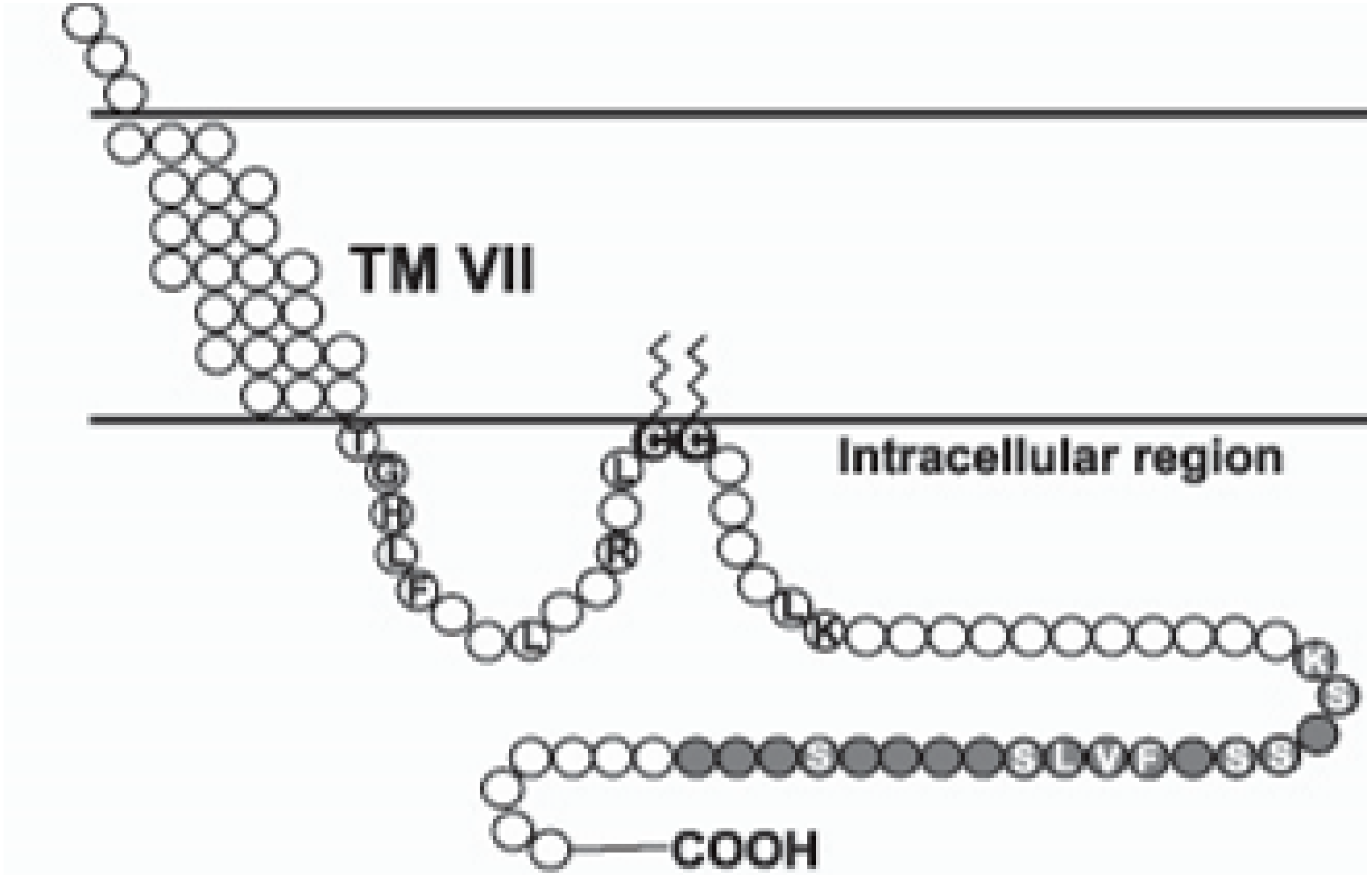

Fig.1

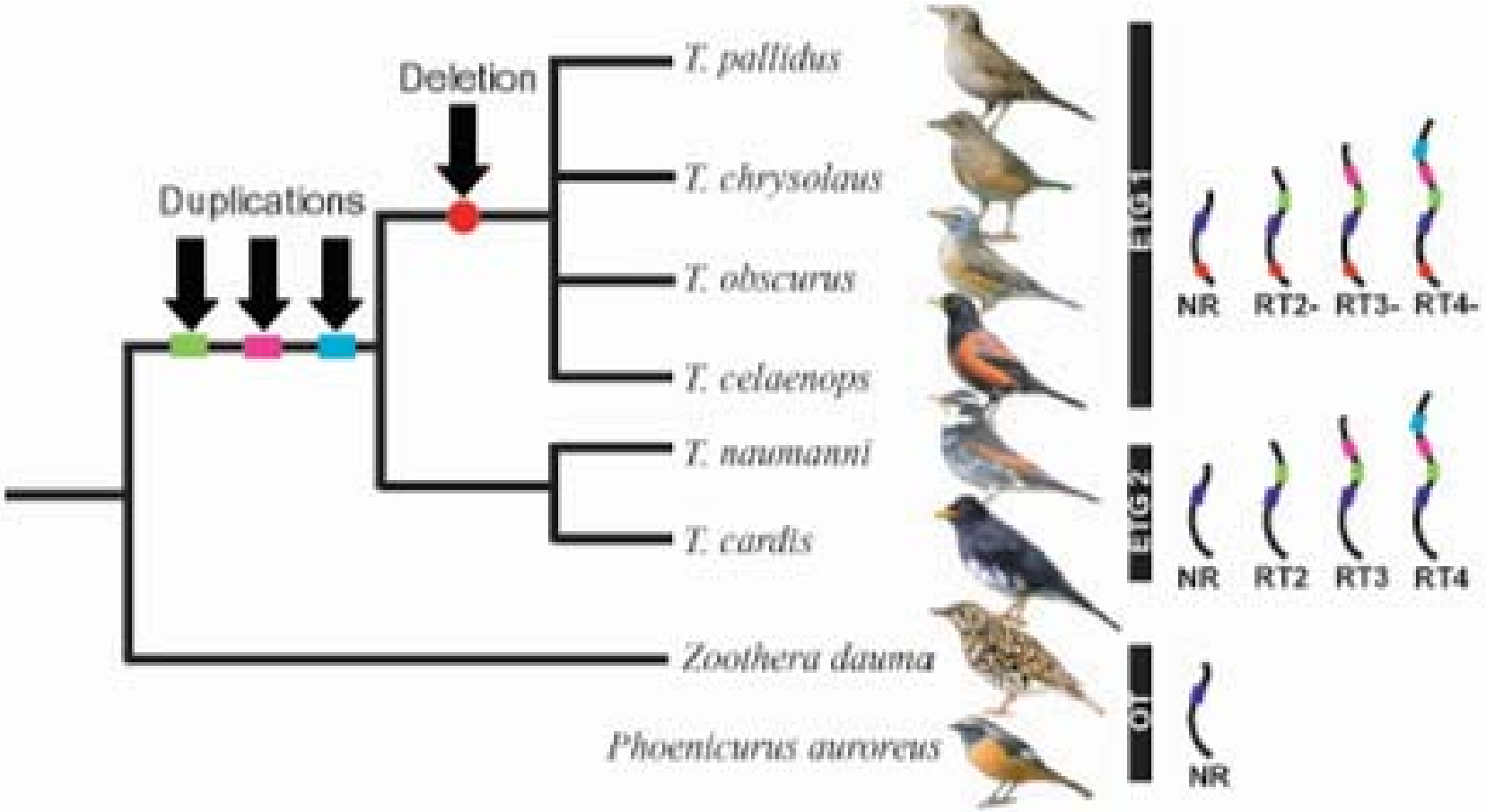
Deletion
Duplications
T. pallidus
T. chrysolaus
T. obscurus
T. celaenops
T. naumanni
T. cardis
Zoothera dauma
Phoenicurus auroreus
NR
RT2-
RT3-
RT4-
NR
RT2
RT3
RT4
NR

Fig. 4

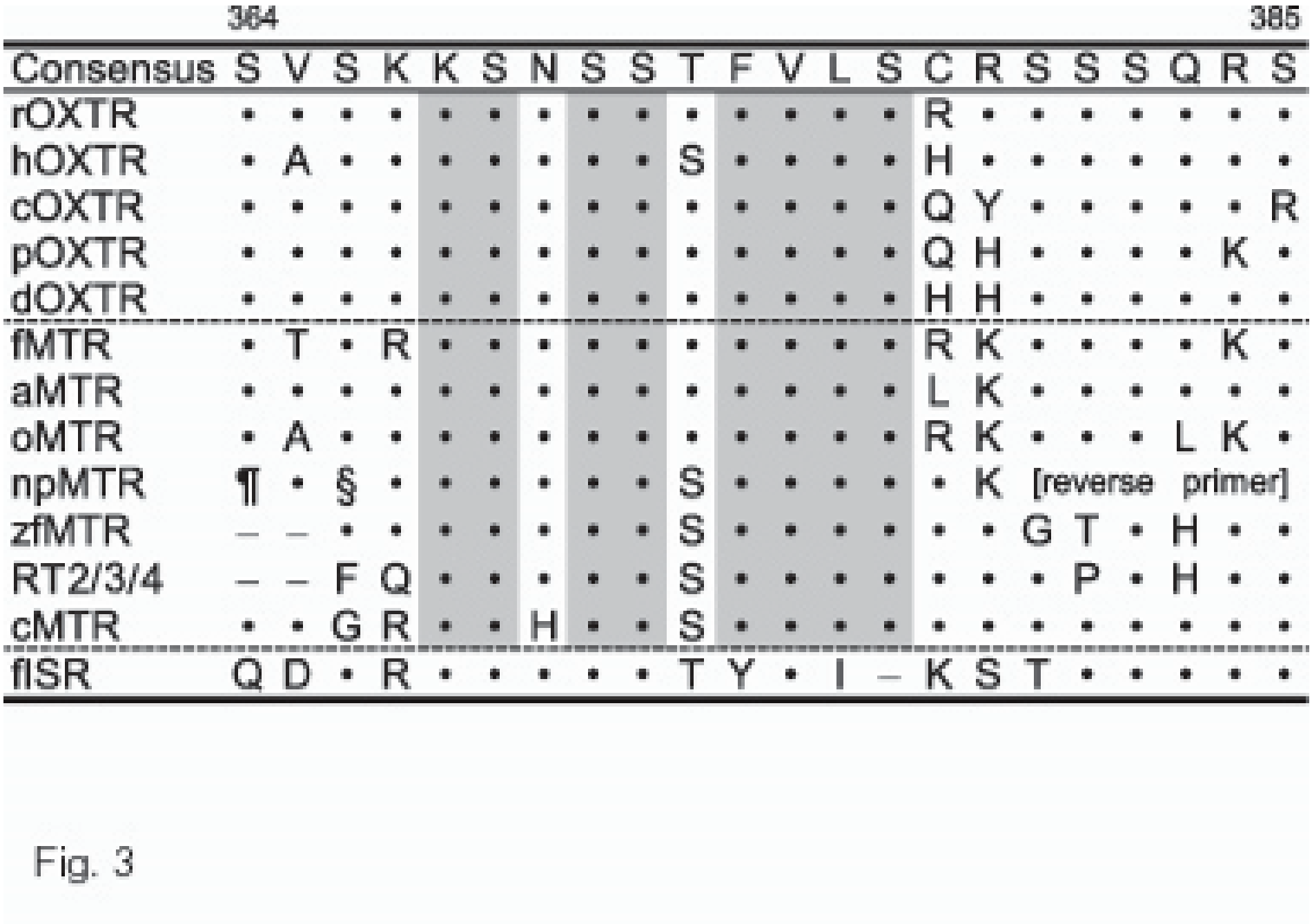
364
385
Consensus S V S K K S N S S T F V L S C R S S S Q R S
rOXTR • • • • • • • • • • • • • • R • • • • • • •
hOXTR • A • • • • • • • S • • • • H • • • • • • •
cOXTR • • • • • • • • • • • • • • Q Y • • • • • R
pOXTR • • • • • • • • • • • • • • Q H • • • • K •
dOXTR • • • • • • • • • • • • • • H H • • • • • •
fMTR • T • R • • • • • • • • • • R K • • • • K •
aMTR • • • • • • • • • • • • • • L K • • • • • •
oMTR • A • • • • • • • • • • • • R K • • • L K •
npMTR ¶ • § • • • • • • S • • • • • K [reverse primer]
zfMTR – – • • • • • • • S • • • • • • G T • H • •
RT2/3/4 – – F Q • • • • • S • • • • • • • P • H • •
cMTR • • G R • • H • • S • • • • • • • • • • • •
flSR Q D • R • • • • • T Y • I – K S T • • • • •

Fig. 3