Supporting Information

Evolutionary renovation of L/M opsin polymorphism confers a fruit discrimination advantage to ateline New World monkeys

Yoshifumi Matsumoto, Chihiro Hiramatsua, Yuka Matsushita, Norihiro Ozawa, Ryuichi Ashino, Makiko Nakata, Satoshi Kasagi, Anthony Di Fiore, Colleen M. Schaffner, Filippo Aureli, Amanda D. Melin and Shoji Kawamura

Contents

Appendix S1: *Inference of ancestral opsins*

At the amino acid site 116, the probability of an inferred residue (Tyr) was less than 0.9 under the JTT model at the node A1 (0.60), A2 (0.65), A3 (0.65). However, the same residue was inferred under the Dayhoff model with high probabilities (0.96, 0.97, and 0.97 at A1, A2, and A3, respectively). At site 225 of node A1, the probability of inference was also less than 0.9 but the same residue (Val) was inferred by both JTT and Dayhof models with probabilities 0.78 and 0.73, respectively. At all the other sites and nodes, probabilities of inferred residues were over 0.9 under both of the substitution models.

Appendix S2: *An alternative ancestral inference at site 217*

In our inference of ancestral sequences, Q217K was estimated to have occurred at branch A and the reverse mutation K217Q was inferred at branch F (Fig. 3). The probability of the amino acid assignment at this site was high (1.00) at all nodes in either the JTT or Dayhoff model. However, because the sequences considered here include alleles that can recombine with each other, and because the models do not consider a possibility of information transfer between OTUs, an alternative scenario is plausible for this site: Q217K occurred at branch B (SYT allele of common ancestor of spider and woolly monkeys) and the woolly monkey AFT allele gained this mutation from SYT allele by recombination at branch G (Fig. S2). Fig. S2 is the modified version of Fig. 3 regarding the amino acid status at site 217. This modification changes ancestral opsins A2, A3 and A4 at site 217. The alternative A2 (designated AA2) is equivalent with A1_S33N/S35A/Y213D in Table 2 (λ_{max} 554 nm) and the alternative A3 (AA3) is equivalent with A2 (λ_{max} 556 nm).

We reconstructed the alternative A4 (AA4) by introducing K217Q to A4 and measured its λ_{max} to be 539 nm (Table S1). Effects of mutations in these AA opsins were also equivalent to those in A opsins: N294K caused a large spectral shift in AA2 (-4 nm), Y277F/N294K caused -16 nm shift and explained the spectral difference between AA2 and AA4 (15 nm), S180A causes only a small spectral shift in AA2 (-1 nm) and in AA4 (+1 nm) (Table S1). Thus, our conclusions based on the mutation scheme depicted in Fig. 3 are not changed by this alternative ancestral inference at site 217.

Appendix S3: *An alternative scenario of ancestral sequence inference using a different out-group*

The bootstrap values in Fig. 2 are low overall, possibly because the evolutionary history differs among nucleotide sites with regards to allelic recombination (or gene conversion between paralogs in the case of howler monkeys), which homogenizes the alleles (or the paralogs) within a species. Furthermore, it is also because balancing selection on allelic difference (or purifying selection against gene conversion) may vary among nucleotide sites (Boissinot *et al.* 1998). Regarding the two alleles of spider monkeys and woolly monkeys, even with non-synonymous sites only (Fig. 2B), the recombination could have reduced the difference between alleles in each species and thereby group them closer to each other than they should be without recombination (e.g. the SYT type sequences of howler monkey, capuchin monkey and marmoset could in fact be an in-group, closer to the SYT alleles of spider and woolly monkeys than to their SFT/AFT alleles).

Thus, instead of using the SYT type sequences of howler monkey, capuchin

monkey and marmoset as an out group, we also tested the human L and M opsin sequences (GenBank Z68193 and AC092402) as a reliable out group (and the mouse M opsin sequence for the phylogenetic root). This is to infer ancestral sequences of the four ateline sequences irrespective of their “true” phylogenetic positions among L/M opsin gene types of New World monkeys.

Regarding the phylogenetic relationship among the four ateline sequences, we retained the assumption of the series of bifurcations as shown in Fig. 3. Even under the low bootstrap probabilities given to the nodes of the four sequences (Fig. 2), the actual steps of phylogenetic differentiation of two alleles from two different species are unlikely to be a star phylogeny where the four sequences originated from a single node, but rather should be a series of bifurcations, where two alleles arose in common ancestor of the two species or in each species independently. Only the former pattern was realized in our phylogenetic analyses as in Fig. 2.

Fig. S3 shows the amino acid substitutions thus inferred under JTT and Dayhoff models in common. Difference of amino acid substitutions from Fig. 3 is indicated in red. At branch A, additional substitutions were inferred due to longer evolutionary distance from the node A2 to the human L and M opsin genes than to the other New World monkey SYT type genes at sites 61, 173, 230, and 275. Two substitutions each were removed from branches C to B at sites 229 and 233 and from branches D to E and F to G at sites 18 and 320 with opposite orientation. Spectral effects of F229I and G233S in branch C of Fig.3 were negligible (Table 2). Spectral effects of these and reverse substitutions were also shown to be negligible in the L/M opsin alleles of squirrel monkey (Hiramatsu *et al.* 2004). The spectral effect of the amino acid substitutions at branches D and F in Fig. 3, which include

N18S and I320V, were also considered negligible because only 3 nm and 1 nm shifts occur in these branches, respectively. Thus, conclusions derived from Fig. 3 are not affected by choice of out-group sequences.

Materials and Methods S1: *PCR and DNA sequencing*

For each of the six coding exons of the L/M opsin gene, primer pairs for polymerase chain reaction (PCR) were set outside of the exon. The primer pairs used were: 5’-gcaaggggtgggaggaggaggtgta-3’ and 5’-ctgctgctggggacgtgaagaggag-3’ for a 396-bp amplicon containing the 112-bp exon 1; 5’-tctatggaagggcagaggactct-3’ and 5’-tgtctccccccaggaaaagcagat-3’ for a 1046-bp amplicon containing the 297-bp exon 2; 5’-cctcctgattttgaaagctgtcagctgg-3’ and 5’-cagaggggcccagagaaaggaatgattt-3’ for a 361-bp amplicon containing the 169-bp exon 3; 5’-actggctgccggcccttctctccag-3’ and 5’-aggccaggagggatcgggggcttac-3’ for a 216-bp amplicon containing the 166-bp exon 4; 5’-tatgcctgggtcacctgcctctt-3’ and 5’-tcagagacacgactccaggtgga-3’ for a 393-bp amplicon containing the 240-bp exon 5; 5’-ttggggaacacacttcaacccag-3’ and 5’-ctctcactgaccatcgtcctctt-3’ for a 985-bp amplicon containing the 111-bp exon 6. PCR was carried out at 95˚C for 5 min followed by 40 cycles of 96˚C for 10 s, 60˚C for 30 s, and 72˚C for 2 min. The exons 2 and 6 were sequenced using primers set just outside of them (exon 2, 5’-tctctcctctgcctcctgcccgcag-3’ and 5’-tgagcctgggccctgactggcttac-3’; exon 6, 5’-gctcaatcactttctgtccttccag-3’ and 5’-ggacgggtaggaggcagacc-3’) and the other exons were sequenced using the same primers used for the PCR. The sequencing reaction was carried out in both strands directly with the amplified DNA fragments using Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems Japan, Tokyo) and the sequences were read

using an Applied Biosystems model 3130xl Genetic Analyzer. The nucleotide sequences were confirmed in duplicated PCRs.

Materials and Methods S2: *Opsin photopigment reconstitution*

Opsin cDNAs were cloned into the pMT5 expression vector (Khorana *et al.* 1988) and were transfected into cultured COS-1 cells (RIKEN Cell Bank, Tsukuba, Japan). Cells were incubated with 5 μM 11-*cis* retinal (Storm Eye Institute, Medical University of South Carolina, Charleston), and were solubilized with 1% dodecyl maltoside. The reconstituted photopigments were purified using immobilized 1D4 antibody (University of British Columbia, Vancouver) as previously described (Kawamura & Yokoyama 1998). The absorption spectra of the photopigments were recorded from 250 to 750 nm at 0.5 nm intervals using the U3010 dual beam spectrometer (Hitachi High-Technologies, Tokyo) at 20ºC. Spectra were read 10 times under dark conditions and 10 more times after 3 min of light exposure with Kodak Wratten Gelatin Filter No.3, which cuts off wavelengths shorter than 440 nm from a 60 watt room lamp. The λ_{max} values were given with ± 1 SD. Recorded spectra were analyzed using UV Solution software (Hitachi High-Technologies, Tokyo).

Materials and Methods S3: *Calculation of JND values*

The noise value was given for each receptor mechanism by incorporating both the effect of quantum catch amount and cone proportion in the retina,

$$e_i = \sqrt{\frac{1}{f_i q_i} + \frac{{w_i}^2}{p_i}}$$

where f_i is quantum catch of receptor *i* averaged for fruit and leaf of the species of interest. The quantum catch was given relative to the quantum catch from a 100% reflective surface, q_i is the estimate of quantum flux (in terms of the number of photons) per cone cell per second in receptor *i*, which has an effect only at low light levels, w_i is the Weber fraction of receptor *i*, and p_i is the relative proportion of receptor *i* to the most abundant cone in the retina (Higham *et al.* 2010). We considered bright light condition, where $q_L,\ q_M = 10^8$ and dim light condition, where $q_L,\ q_M = 10^4$ (Land & Nilsson 2002). We set the value of quantum flux for S cones as $q_s = q_L/10$ by taking account of low sensitivity of S cones. As Weber fractions, we set $w_L,\ w_M = 0.02,\ w_s = 0.08$. Those values are close to psychophysical thresholds for humans (Osorio *et al.* 2004; Wyszecki & Stiles 1982). Based on the cone proportion in marmoset (Martin & Grunert 1999), we set $p_L,$ $p_M = 1,\ p_s = 0.1$. Then the chromatic JND value for trichromats was calculated by

$$\mathrm{JND} = \sqrt{\frac{{e_S}^2(\Delta f_L - \Delta f_M)^2 + {e_M}^2(\Delta f_L - \Delta f_S)^2 + {e_L}^2(\Delta f_M - \Delta f_S)^2}{(e_L e_M)^2 + (e_L e_S)^2 + (e_M e_S)^2}}$$

where Δf_i is difference of quantum catch of receptor *i* between from a fruit and from its background leaf and is given by,

$$\Delta f_i = \ln\frac{f_{i1}}{f_{i2}}$$

where f_{i1} and f_{i2} are quantum catch of receptor *i* from a fruits and its leaf, respectively (Osorio *et al.* 2004).

Table S1 λ_{max} of opsins with introduced mutations under an alternative ancestral inference at site 217

Opsin	λ_{max}±SE (nm)
AA2 (= A1_S33N/S35A/Y213D; Table 2)	554±1.3
AA3 (= A2; Table 2)	556±2.0
AA4	539±1.0
AA2_N294K	550±0.9
AA2_Y277F/N294K (= A2_Y277F/N294K/K217Q; Table 2)	538±2.2
AA2_S180A	553±0.3
AA4_S180A	540±1.5

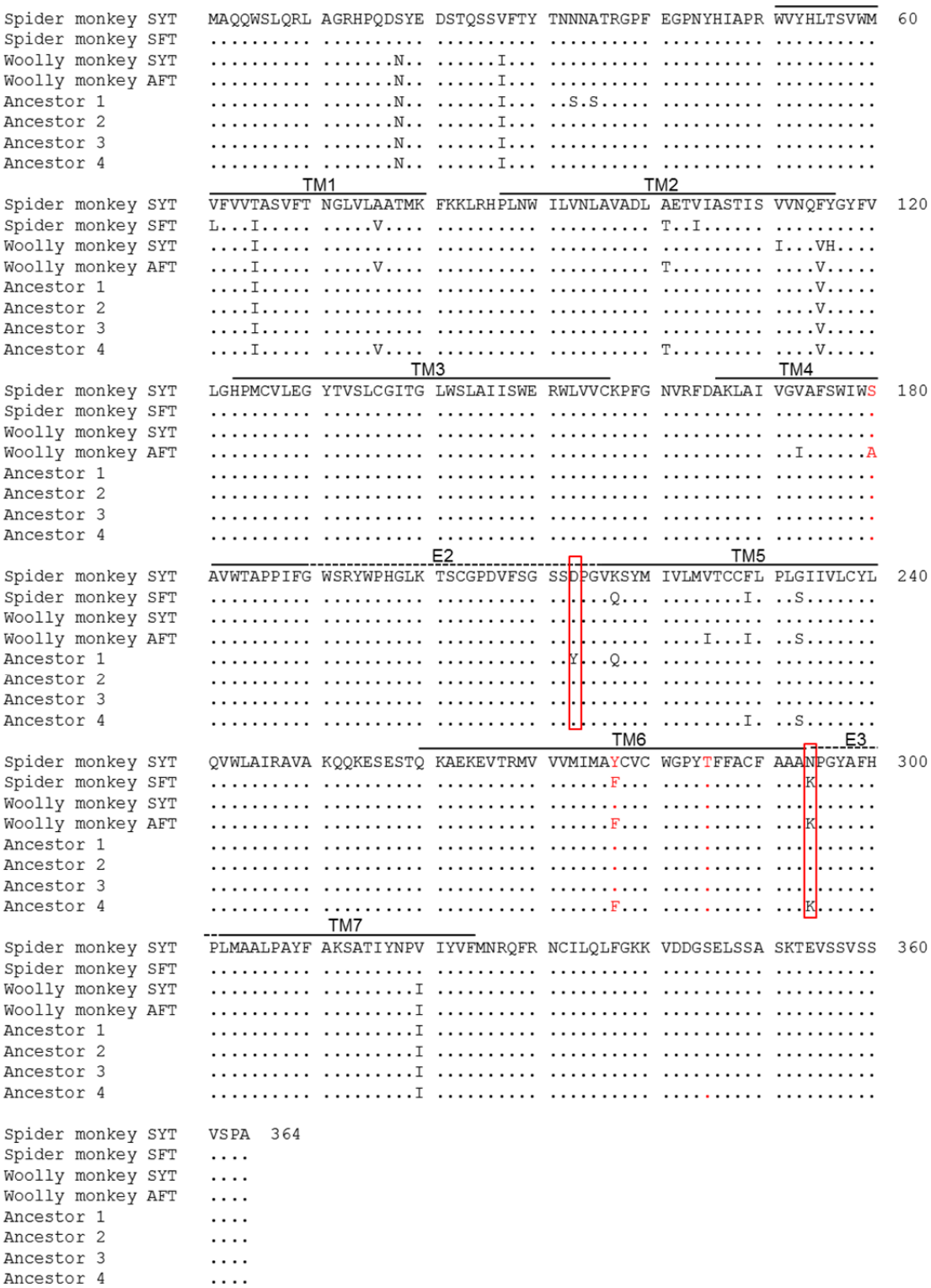

Fig. S1 Amino acid sequences of the L/M opsin types found in spider and woolly monkeys and their estimated ancestral opsins. The complete sequence of the spider

monkey SYT opsin is provided. Identical amino acids to the spider monkey SYT are indicated by dots. The seven TM regions and the second and third extracellular loop regions (E2 and E3) are indicated with solid and dotted lines, respectively. The 180, 277 and 285 amino acid sites are depicted in red. The 213 and 294 amino acid sites are boxed.

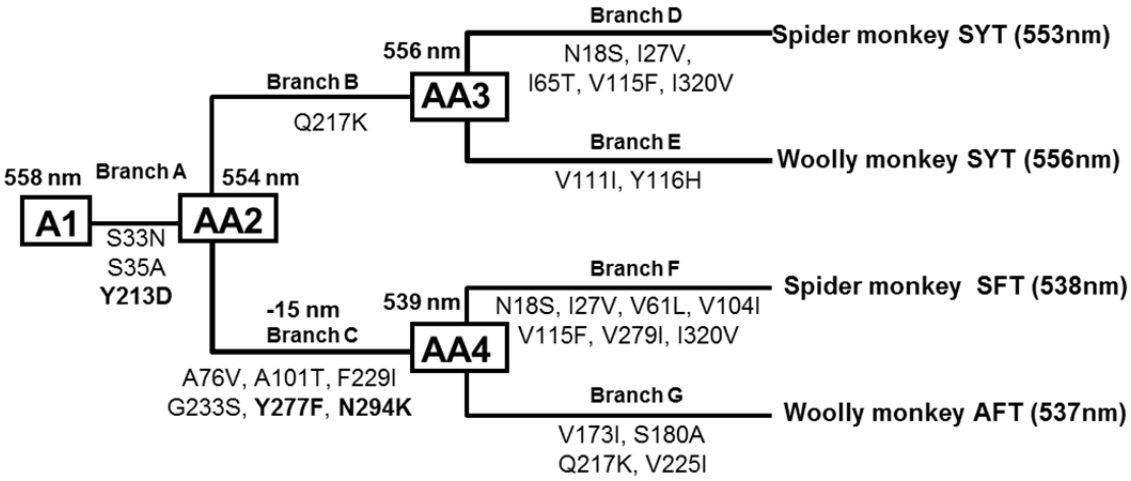

Fig. S2 An alternative ancestral inference at site 217. Figure legend follows that of Fig. 3. Assignment of Q217K is different from Fig. 3. The alternative ancestral opsins 2~4 (AA2~AA4) and their λ_{max} are indicated.

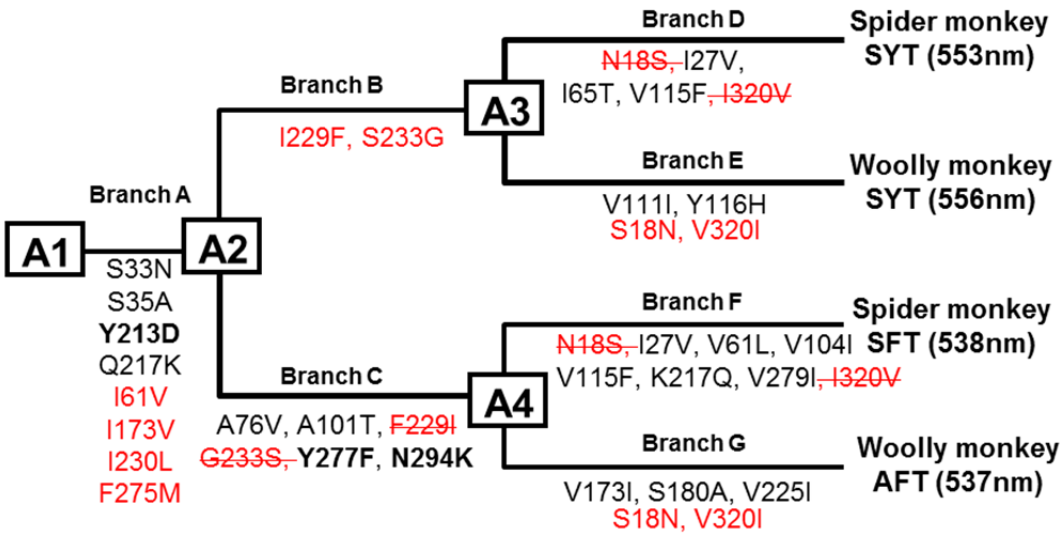

Fig. S3 An alternative scenario of ancestral sequence inference by different out-group. Figure legend follows that of Fig. 3. Additional mutations to Fig. 3 are indicated with red letters. Fig. 3 mutations not supported here are erased with a bar and indicated with red color.

References

Boissinot S, Tan Y, Shyue SK, *et al.* (1998) Origins and antiquity of X-linked triallelic color vision systems in New World monkeys. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 13749-13754.

Higham JP, Brent LJ, Dubuc C, *et al.* (2010) Color signal information content and the eye of the beholder: a case study in the rhesus macaque. *Behavioral ecology : official journal of the International Society for Behavioral Ecology* **21**, 739-746.

Hiramatsu C, Radlwimmer FB, Yokoyama S, Kawamura S (2004) Mutagenesis and reconstitution of middle-to-long-wave-sensitive visual pigments of New World monkeys for testing the tuning effect of residues at sites 229 and 233. *Vision Research* **44**, 2225-2231.

Kawamura S, Yokoyama S (1998) Functional characterization of visual and nonvisual pigments of American chameleon (*Anolis carolinensis*). *Vision Research* **38**, 37-44.

Khorana HG, Knox BE, Nasi E, Swanson R, Thompson DA (1988) Expression of a bovine rhodopsin gene in *Xenopus* oocytes: demonstration of light-dependent ionic currents. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 7917-7921.

Land MF, Nilsson D-E (2002) *Animal Eyes* Oxford University Press, Oxford.

Martin PR, Grunert U (1999) Analysis of the short wavelength-sensitive ("blue") cone mosaic in the primate retina: comparison of New World and Old World monkeys. *The Journal of Comparative Neurology* **406**, 1-14.

Osorio D, Smith AC, Vorobyev M, Buchanan-Smith HM (2004) Detection of fruit and the selection of primate visual pigments for color vision. *The American Naturalist* **164**, 696-708.

Wyszecki G, Stiles WS (1982) *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd edn. Wiley, New York.