*Supplementary Material for*

**The rice paradox: Multiple origins but single domestication in Asian rice**

Jae Young Choi1, Adrian E. Platts1, Dorian Q. Fuller2, Yue-Ie Hsing (邢禹依)3, Rod A. Wing4 and Michael D. Purugganan1,5

1Center for Genomics and Systems Biology, Department of Biology, 12 Waverly Place, New York University, New York, NY 10003 USA

2Institute of Archaeology, University College London, 31-34 Gordon Square, London WC1H 0PY, United Kingdom

3Institute of Plant and Microbial Biology, Academica Sinica, 128 Sec. 2, Academia Rd, Nankang, Taipei 11529 Taiwan, Republic of China

4Arizona Genomics Institute, University of Arizona, 1657 E Helen St, Tucson, AZ 85705 USA

5Center for Genomics and Systems Biology, New York University Abu Dhabi, Saadiyat Island, Abu Dhabi, United Arab Emirates

**Corresponding Author**: Michael Purugganan. Email: mp132@nyu.edu. Tel. (212) 9983801

Index

Supplementary Text1

Comparing G-PhoCS estimated divergence time in single- vs. multi-migration model1

Supplementary Figures4

Supplementary Tables11

References19

# Supplementary Text

**Comparing G-PhoCS estimated divergence time in single- vs. multi-migration model**

Since our final migration model did not detect any evidence of bidirectional gene flow, we compared divergence time (τ) estimates in the final combined unidirectional migration model, which included all significant migration events, to those that forced bidirectional gene flow only between two terminal lineages (Supplementary Fig. S5). As expected, bidirectional migration models that had no significant evidence of gene flow (for example gene flow between *japonica* and *O. rufipogon*) had τ estimates that were not significantly different from a no migration model. On the other hand, two lineages with significant bidirectional gene flow in Supplementary Fig. S1 had estimates of τ values that were significantly different from the no migration model. Hence, we focused on examining models with significant evidence of bidirectional gene flow (Supplemental Fig. S1).

The models forcing bidirectional gene flow only between *japonica*-*indica* (J↔I) or only between *japonica*-*aus* (J↔A)had *O. rufipogon* – *japonica* divergence time (τRJ)estimates that were similar to the final combined unidirectional model. This suggested that the mis-incorporation of the *aus*/*indica* to *japonica* gene flow did not strongly affect estimates of τRJ. The divergence time for domesticated and wild rice population (τroot) estimate for the J↔I and J↔A models were significantly greater then the no migration model, but were still significantly lower then the final combined unidirectional model. However, the J↔I and J↔A models each underestimates the total number of migration events and this can lead to an underestimate of τ (Gronau et al. 2011).

Meanwhile, compared to the J↔I and J↔A models, the *aus*/*O. nivara* split time (τAN) and *indica* split time (τIAN) were overestimated in the final combined unidirectional model. However, it is unclear if the individual bidirectional gene flow models are estimating the correct τ for lineages involving *aus, indica,* and *O. nivara*, as the model forcing a potentially non-existent gene flow between *japonica*-*O. nivara* (J↔N) resulted in similar τAN and τIAN estimates as the J↔I and J↔A models.

In comparison, the *aus*-*indica* bidirectional gene flow (A↔I) model also estimated similar τAN value to the J↔I and J↔A models; however, the estimate of τIAN was almost as large as the τRJ estimates which is not supported by the archaeological data (Fuller 2011). Thus, compared to the individual bidirectional gene flow models, the higher τAN and τIAN in the final combined unidirectional gene flow model may represent an upper bound τ estimates. These results suggested that the τ estimates in the final combined unidirectional gene flow model were not greatly affected by omitting the gene flows detected from the individual bidirectional gene flow models. Importantly the τRJ estimate, which is of interest for this study and rice domestication researchers due to *japonica* being the first cultivar to be domesticated, did not seem to be greatly affected by mis-incorporation of the *aus*/*indica* to *japonica* gene flow. At most the divergence time estimates in the final unidirectional gene flow model may represent an upper bound estimate specifically for τroot, τAN,and τIAN.

# Supplementary Figures



Supplemental Fig. S1. G-PhoCS estimated migration rates and its 95% Highest Posterior Density for domesticated and wild Asian rice. Population are abbreviated as J: japonica; R: *O. ruifipogon*; I: indica; A: aus; N: *O. nivara*. Arrow indicates the direction of gene flow used for fitting the demography model (source population→target population). G-PhoCS analysis were conducted for 5 independent runs to access convergence in the parameter estimates. Parameter estimates have been scaled down by a factor of 10-4 for ease of representing.



Supplemental Fig. S2. G-PhoCS estimated migration rates and its 95% Highest Posterior Density for domesticated and wild Asian rice. Population are abbreviated as J: japonica; R: *O. ruifipogon*; I: indica; A: aus; N: *O. nivara*. Arrow indicates the direction of gene flow used for fitting the demography model (source population→target population). G-PhoCS analysis were conducted for 5 independent runs to access convergence in the parameter estimates. Migration model 1 was a G-PhoCS analysis modeling without an indica-aus migration band while migration model 2 was modeling with the indica-aus migration band. Parameter estimates have been scaled down by a factor of 10-4 for ease of representing.



Supplemental Fig. S3. G-PhoCS estimated population size (θ) and its 95% Highest Posterior Density for current domesticated and wild Asian rice. Population are abbreviated as J: japonica; R: *O. ruifipogon*; I: indica; A: aus; N: *O. nivara*. G-PhoCS analysis were conducted for 5 independent runs to access convergence in the parameter estimates. θ was estimated under two different scenarios: with and without migration bands. Parameter estimates have been scaled up by a factor of 104 for ease of representing.



Supplemental Fig. S4. G-PhoCS estimated divergence time (τ) and its 95% Highest Posterior Density for current domesticated and wild Asian rice. Population are abbreviated as J: japonica; R: *O. ruifipogon*; I: indica; A: aus; N: *O. nivara*. G-PhoCS analysis were conducted for 5 independent runs to access convergence in the parameter estimates. τ was estimated under two different scenarios: with and without migration bands. Parameter estimates have been scaled up by a factor of 104 for ease of representing.



Supplemental Fig. S5. G-PhoCS estimated divergence time (τ) and its 95% Highest Posterior Density under various gene flow model. Population are abbreviated as J: japonica; R: *O. ruifipogon*; I: indica; A: aus; N: *O. nivara*. Bidirectional arrow indicates bidirectional gene flow between two lineages. Mig. Model 2 indicate the migration model 2 in Supplementary Fig S2. Final Mig. Model indicates the final unidirectional gene flow model. No Mig. indicates the no migration model. G-PhoCS analyses were conducted for 5 independent runs to access convergence in the parameter estimates. Parameter estimates have been scaled up by a factor of 104 for ease of representing.



Supplemental Fig. S6. DFOIL test of evaluating direction of introgression. The five-taxon phylogeny represents the topology that is required for the DFOIL test. The DFOIL test is a combination of four ABBA-BABA test like D-statistics (DFO, DIL, DFI, and DOL). Each color represents an ABBA-BABA test comparison involving two closely related monophyletic group and a third ingroup lineage. Signs (+ or -) for each DFOIL component depend on the extent of allele sharing between the third ingroup and one of the two monophyletic sister lineages.

# Supplementary Tables

Supplemental Table S1. Genome alignment statistic using the japonica genome as reference.

|  |  |
| --- | --- |
| Target Genome | aCoverage % |
| ausDJ123 | 74.0% |
| ausKasalath | 83.2% |
| indica93-11 | 71.4% |
| indicaIR64 | 74.2% |
| *O. nivara* | 65.1% |
| *O. rufipogon* | 74.3% |
| *O. punctata* | 36.9% |

a. Proportion of the reference japonica genome that had a target genome’s base aligned to it.

Supplemental Table S2. Total number of genes analyzed for the four-taxon topology test. *O. punctata* was always the outgroup species

|  |  |
| --- | --- |
| Compared Population | Number ofAnalyzed genes |
| ausDJ123, *O.nivara, O.rufipogon* | 14346 |
| ausKasalath, *O.nivara, O.rufipogon* | 17965 |
| indicaIR64, *O.nivara, O.rufipogon* | 19129 |
| indicaIR64, ausDJ123, *O.nivara* | 19295 |
| indicaIR64, ausKasalath, *O.nivara* | 18139 |
| indica93-11, *O.nivara, O.rufipogon* | 18442 |
| indica93-11, ausDJ123, *O.nivara* | 18670 |
| Indica93-11, ausKasalath, *O. nivara* | 17617 |
| japonica, *O.nivara, O.rufipogon* | 19707 |
| japonica, ausDJ123, *O.nivara* | 19857 |
| japonica, ausDJ123, *O.rufipogon* | 21136 |
| japonica, ausKasalath, *O.nivara* | 18670 |
| japonica, ausKasalath, *O.rufipogon* | 19478 |
| japonica, indicaIR64, *O.nivara* | 19903 |
| japonica, indicaIR64, *O.rufipogon* | 21187 |
| japonica, indicaIR64, ausDJ123 | 21405 |
| japonica, indicaIR64, ausKasalath | 20069 |
| japonica, indica93-11, *O.nivara* | 19287 |
| japonica, indica93-11, *O.rufipogon* | 20493 |
| japonica, indica93-11, ausDJ123 | 20702 |
| japonica, indica93-11, ausKasalath | 19478 |

Supplemental Table S3. Total number of gene trees significantly supporting a topology after the Approximately Unbiased (AU) test. Numbers in parenthesis represent percentage with 95% bootstrap confidence interval indicated in square brackets. *O. punctata* was used as outgroup for all topology tests.

|  |  |
| --- | --- |
| **Major Topology** | **Minor Topologies** |
| ((japonica,O.rufipogon),O.nivara) | ((japonica,O.nivara),O.rufipogon) | ((O.nivara,O.rufipogon),japonica) |
| 5409 (85.1% [84.3-86.0%]) | 497 (7.8% [7.2-8.5%]) | 446 (7.0% [6.4-7.7%]) |
| ((SAADJ123,O.nivara),O.rufipogon) | ((SAADJ123,O.rufipogon),O.nivara) | ((O.nivara,O.rufipogon),SAADJ123) |
| 2214 (65.0% [63.4-66.6%]) | 975(28.6% [27.1-30.1%]) | 215 (6.3% [5.5-7.1%]) |
| ((SAAKasalath,O.nivara),O.rufipogon) | ((SAAKasalath,O.rufipogon),O.nivara) | ((O.nivara,O.rufipogon),SAAKasalath) |
| 2871 (66.1% [64.7-67.5%]) | 1154 (26.5% [25.3-27.9%]) | 319 (7.3% [6.6-8.1%]) |
| ((indica93-11,O.nivara),O.rufipogon) | ((indica93-11,O.rufipogon),O.nivara) | ((O.nivara,O.rufipogon),indica93-11) |
| 2418 (54.9% [53.4-56.3%]) | 1570 (35.6% [34.2-37.0%]) | 418 (9.5% [8.6-10.3%]) |
| ((indicaIR64,O.nivara),O.rufipogon) | ((indicaIR64,O.rufipogon),O.nivara) | ((O.nivara,O.rufipogon),indicaIR64) |
| 2499 (56.9% [55.5-58.4%]) | 1447 (33.0% [31.6-34.3%]) | 444 (10.1% [9.2-11.0%]) |
| ((SAADJ123,O.nivara),indicaIR64) | ((SAADJ123,indicaIR64),O.nivara) | ((indicaIR64,O.nivara),SAADJ123) |
| 2854 (58.8% [57.4-60.3%]) | 1113 (22.9% [21.7-24.1%]) | 883 (18.2 [17.1-19.3%]) |
| ((SAAKasalath,O.nivara),indicaIR64) | ((SAAKasalath,indicaIR64),O.nivara) | ((indicaIR64,O.nivara),SAAKasalath) |
| 2662 (59.9% [58.5-61.4%]) | 1008 (22.7% [21.4-23.9%]) | 771 (17.4% [16.2-18.5%]) |
| ((SAADJ123,O.nivara),indica93-11) | ((SAADJ123,indica93-11),O.nivara) | ((indica93-11,O.nivara),SAADJ123) |
| 2726 (58.3% [56.9-59.8%]) | 1117 (23.9% [22.7-251.%]) | 831 (17.8% [16.7-18.9%]) |
| ((SAAKasalath,O.nivara),indica93-11) | ((SAAKasalath,indica93-11),O.nivara) | ((indica93-11,O.nivara),SAAKasalath) |
| 2511 (58.0% [56.5-59.4%]) | 1061 (24.5% [23.2-25.8%]) | 759 (17.5% [16.4-18.7%]) |
| ((SAADJ123,O.nivara),japonica) | ((japonica,SAADJ123),O.nivara) | ((japonica,O.nivara),SAADJ123) |
| 3062 (67.1% [65.8-68.5%]) | 965 (21.1% [20.0-22.4%]) | 535 (11.7% [10.8-12.7%]) |
| ((japonica,O.rufipogon),SAADJ123) | ((japonica,SAADJ123),O.rufipogon) | ((SAADJ123,O.rufipogon),japonica) |
| 4887 (77.4% [76.4-78.4%]) | 972 (15.4% [14.5-16.3%]) | 451 (7.1% [6.5-7.8%]) |
| ((SAAKasalath,O.nivara),japonica) | ((japonica,SAAKasalath),O.nivara) | ((japonica,O.nivara),SAAKasalath) |
| 2912 (68.2% [66.8-69.7%]) | 849 (19.9% [18.7-21.1%]) | 505 (11.8% [10.9-12.8%]) |
| ((japonica,O.rufipogon),SAAKasalath) | ((japonica,SAAKasalath),O.rufipogon) | ((SAAKasalath,O.rufipogon),japonica) |
| 4540 (78.5% [77.4-79.6%]) | 839 (14.5% [13.6-15.4%]) | 405 (7.0% [6.3-7.7%]) |
| ((indicaIR64,O.nivara),japonica) | ((japonica,indicaIR64),O.nivara) | ((japonica,O.nivara),indicaIR64) |
| 2562 (59.9% [58.4-61.3%]) | 1025 (23.9% [22.7-25.2%]) | 693 (16.2% [15.1-17.3%]) |
| ((japonica,O.rufipogon),indicaIR64) | ((japonica,indicaIR64),O.rufipogon) | ((indicaIR64,O.rufipogon),japonica) |
| 4932 (77.9% [76.9-79.0%]) | 947 (15.0% [14.1-15.8%]) | 450 (7.1% [6.5-7.8%]) |
| ((indica93-11,O.nivara),japonica) | ((japonica,indica93-11),O.nivara) | ((japonica,O.nivara),indica93-11) |
| 2480 (58.5% [57.0-60.0%]) | 1143 (27.0% [25.6-28.3%]) | 617 (14.6% [13.5-15.6%]) |
| ((japonica,O.rufipogon),indica93-11) | ((japonica,indica93-11),O.rufipogon) | ((indica93-11,O.rufipogon),japonica) |
| 4499 (75.6% [74.5-76.7%]) | 1034 (17.4% [16.4-18.4%]) | 417 (7.0% [6.4-7.6%]) |

Supplemental Table S4. ABBA-BABA test for four populations (P1,P2,P3,O). The *O. punctata* genome was used as the outgroup genome (O).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| P1 | P2 | P3 | ABBA sites | BABA sites | D | SE | Z-score |
| *O. rufipogon* | japonica | ausDJ123 | 55873 | 31422 | 0.2801 | 0.0516 | **5.42829** |
| indica93-11 | indicaIR64 | ausDJ123 | 36441 | 35905 | 0.00741 | 0.06968 | 0.10634 |
| *O. rufipogon* | japonica | ausKasalath | 52744 | 29107 | 0.28878 | 0.04528 | **6.37765** |
| indica93-11 | indicaIR64 | ausKasalath | 32659 | 36331 | -0.05323 | 0.06033 | -0.88231 |
| *O. nivara* | ausDJ123 | indica93-11 | 58930 | 53599 | 0.04737 | 0.06436 | 0.73602 |
| *O. nivara* | ausKasalath | indica93-11 | 58414 | 46286 | 0.11584 | 0.05519 | 2.09893 |
| *O. rufipogon* | japonica | indica93-11 | 58892 | 30055 | 0.3242 | 0.04235 | **7.65525** |
| ausDJ123 | ausKasalath | indica93-11 | 36579 | 30143 | 0.09646 | 0.07023 | 1.37349 |
| *O. nivara* | ausDJ123 | indicaIR64 | 58370 | 54617 | 0.03322 | 0.05518 | 0.60203 |
| *O. nivara* | ausKasalath | indicaIR64 | 55395 | 49034 | 0.06091 | 0.05255 | 1.15909 |
| *O. rufipogon* | japonica | indicaIR64 | 55955 | 31426 | 0.28071 | 0.04843 | **5.7962** |
| ausDJ123 | ausKasalath | indicaIR64 | 34508 | 31455 | 0.04628 | 0.07114 | 0.65055 |
| *O. nivara* | ausDJ123 | japonica | 65702 | 27461 | 0.41047 | 0.06671 | **6.15305** |
| *O. nivara* | ausKasalath | japonica | 60243 | 24190 | 0.427 | 0.05159 | **8.2768** |
| *O. nivara* | indicaIR64 | japonica | 71212 | 36520 | 0.32202 | 0.05492 | **5.86344** |
| *O. nivara* | indica93-11 | japonica | 79451 | 32779 | 0.41586 | 0.04548 | **9.1438** |
| ausDJ123 | ausKasalath | japonica | 26613 | 28951 | -0.04208 | 0.11031 | -0.38147 |
| indicaIR64 | ausDJ123 | japonica | 57265 | 53929 | 0.03 | 0.07243 | 0.41419 |
| indicaIR64 | ausKasalath | japonica | 54095 | 51503 | 0.02455 | 0.07005 | 0.35046 |
| indica93-11 | ausDJ123 | japonica | 52952 | 61387 | -0.07377 | 0.06673 | -1.1055 |
| indica93-11 | ausKasalath | japonica | 47703 | 55884 | -0.07898 | 0.06418 | -1.2306 |
| indica93-11 | indicaIR64 | japonica | 26063 | 37876 | -0.18475 | 0.06703 | -2.75623 |
| *O. rufipogon* | japonica | *O. nivara* | 43597 | 35546 | 0.10173 | 0.05489 | 1.85334 |
| ausDJ123 | ausKasalath | *O. nivara* | 37340 | 32805 | 0.06465 | 0.09308 | 0.69456 |
| indica93-11 | indicaIR64 | *O. nivara* | 36420 | 34373 | 0.02892 | 0.05146 | 0.56199 |
| *O. nivara* | ausDJ123 | *O. rufipogon* | 54327 | 32606 | 0.24986 | 0.05301 | **4.71345** |
| *O. nivara* | ausKasalath | *O. rufipogon* | 50974 | 29685 | 0.26394 | 0.05836 | **4.52262** |
| *O. nivara* | indicaIR64 | *O. rufipogon* | 60654 | 42094 | 0.18064 | 0.05313 | 3.39996 |
| *O. nivara* | indica93-11 | *O. rufipogon* | 65547 | 39474 | 0.24826 | 0.04963 | **5.00222** |
| ausDJ123 | ausKasalath | *O. rufipogon* | 25242 | 26479 | -0.02392 | 0.0732 | -0.32678 |
| indicaIR64 | ausDJ123 | *O. rufipogon* | 53045 | 49658 | 0.03298 | 0.05894 | 0.55955 |
| indicaIR64 | ausKasalath | *O. rufipogon* | 50856 | 47327 | 0.03594 | 0.0592 | 0.60709 |
| indica93-11 | ausDJ123 | *O. rufipogon* | 50463 | 54537 | -0.0388 | 0.05606 | -0.69212 |
| indica93-11 | ausKasalath | *O. rufipogon* | 45862 | 49371 | -0.03685 | 0.05754 | -0.64042 |
| indica93-11 | indicaIR64 | *O. rufipogon* | 25885 | 33613 | -0.12989 | 0.06066 | -2.14128 |

Z-score higher then 3.9 (p < 0.0001) are bolded.

Supplemental Table S5. DFOIL test statistic and corresponding chi-squared values. The *O. punctata* genome was used as the outgroup genome (O).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Compared Population (P1,P2,P3,P4) | DFO | DIL | DFI | DOL |
| Statistic | Χ2 | Statistic | Χ2 | Statistic | Χ2 | Statistic | Χ2 |
| *O. nivara,*ausDJ123,japonica,*O. rufipogon* | **0.13** | 5230.30 | **0.24** | 17547.66 | **-0.28** | 26451.56 | **-0.18** | 11113.21 |
| *O. nivara,*ausKasalath,japonica,*O. rufipogon* | **0.14** | 5401.44 | **0.24** | 16448.29 | **-0.25** | 19959.73 | **-0.16** | 7917.26 |
| *O. nivara,*indicaIR64,japonica,*O. rufipogon* | **0.13** | 5250.03 | **0.24** | 17204.83 | **-0.22** | 18928.69 | **-0.14** | 7484.21 |
| *O. nivara,*indica93-11,japonica,*O. rufipogon* | **0.13** | 5238.57 | **0.27** | 21725.35 | **-0.22** | 19728.12 | **-0.12** | 5945.24 |

D-statistics with chi-squared values higher then 15.14 (p < 0.0001) are bolded.

Supplemental Table S6. Combination of signs for the DFOIL test and its interpretation. Species topology and name can be seen in Supplementary Fig. S6.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Introgression Direction | DFO | DIL | DFI | DOL |
| None | 0 | 0 | 0 | 0 |
| P1→P3 | + | + | + | - |
| P3→P1 | + | 0 | + | + |
| P1→P4 | - | - | 0 | + |
| P4→P1 | - | 0 | + | + |
| P2→P3 | + | + | - | 0 |
| P3→P2 | 0 | + | - | - |
| P2→P4 | - | - | 0 | - |
| P4→P2 | 0 | - | - | - |
| P12→P3 | + | + | 0 | 0 |
| P3→P12 | + | + | 0 | 0 |
| P12→P4 | - | - | 0 | 0 |
| P4→P12 | - | - | 0 | 0 |

# References

Gronau I, Hubisz MJ, Gulko B, Danko CG, Siepel A. 2011. Bayesian inference of ancient human demography from individual genome sequences. Nat. Genet. 43:1031–1034.

Fuller DQ. 2011. Pathways to Asian Civilizations: Tracing the Origins and Spread of Rice and Rice Cultures. Rice 4:78–92.