

Editorial suggestions on Manuscript ID: MER-14-0343 “Interpreting the FLOCK algorithm from a statistical perspective”

Page and Line numbers are from the pdf file.

Page 2, Abstract. Line 18. See “...no-admxixture...” **Spelling**

P. 2, Line 19. “uncorrelated allele frequency prior **albeit** with more variability...”

P. 3, Line 24. “The best known ~~example~~ of this class ...”

P. 3, Lines 44-45. “~~Being able to interpret~~ **Recognizing FLOCK’s** ~~in this fashion~~ **underpinnings** should help users ... when it might ~~and might not~~ give substantially...”

P. 5. Line 68. “...to randomly drawn values **from the prior**, ...”

P. 5, Lines 70-71. “...at iteration $t = 0, 1, 2, \dots, T$,
~~↳~~ each ...”

P. 6, Line 74. “for all $k = 1, \dots, K$, **implying the belief that the subpopulations are equally represented in the sample before observing the genotypes present.**”

Page 7, Lines 107-108. “...the program randomly allocates or ~~each of~~ **partitions** the N individuals **into (approximately) one of K equal-sized** clusters.”

Page 7, Lines 109- ...each individual is ~~given the chance to be reallocated (i.e., moved to a different cluster). This is done on the basis of maximum likelihood:~~ **non-randomly assigned to the cluster for which its multilocus likelihood is maximal:** as Duchesne & Turgeon say ...”

Page 8, Line 138. “If, all that was desired...” Delete comma.

Page 9, Line 155. “... we can expect that it ~~may be more~~ is susceptible to ...”

Page 9, Line 158. “...the same solution **among repeat applications to the same data set** less consistently than does STRUCTURE.”

Page 10, Line 161. “...to compare the programs FLOCK and STRUCTURE. **[add] The criteria for comparison include amount of computer time used, estimates of K , and the allocation of individuals among clusters.**” Now move Lines 177 – 182 to immediately follow, i.e., “**For FLOCK**, the user must specify ... is calculated for each individual.” **Continue with Lines 161 -168, i.e.,** “FLOCK was run six times ... running 64-bit Windows 7.” You might also want to explain the choice of 50,000 sweep burn-in and 150,000 sweep sample.

Page 10. Lines 169 – 175. No change except Page 10, Line 175. “DISTRICT” **spelling**

Page 10, Lines 176 – 198. Omit Lines 177- 182 that have been moved above.

Page 11, Line 197. "...for estimating K with FLOCK to results obtained from STRUCTURE using $\ln P(D)$..."

Page 12, Lines 209 – 228. Figure 1 is referenced which uses lower case k , while this text uses the upper case K .

Page 12, Line 217. "Both programs identify ..." Spelling

Page 12, Lines 230 – 233. "To investigate the stability of the partition arrangement of individuals within from the reference populations to move around the partition space, we varied ... Unfortunately Spelling only the ...is given as FLOCK output."

Page 13, Line 234. "...during the last-later iterations of runs, individuals ..."

Page 14, Lines 247- 248. "...give similar results, only although at values of $K > 4$, ... appear to be more variable among random starting partitions".

Page 14, Lines 260 – 262. "This, coupled with ...of length zero, complicates ... from FLOCK. if no non-zero plateaus are encountered.

Page 15, Lines 270 – 271. "... in their comparisons, we did not observe this advantage, probably for in our analysis. This appears to result from two issues reasons. "

Page 16, Lines 302 – 303. "This is especially true in real populations and the actual conditions under which samples are drawn which may not conform to the assumptions of genetic clustering models."

Page 16, Lines 304 – 307. "We made our comparison between FLOCK and the no-admixture model in STRUCTURE which is not the instead of its default with-admixture model. used in STRUCTURE. Rather, by default, STRUCTURE uses its with-admixture model. Such a The default model explicitly tries to account for the admixed origin of individuals in the sample and thus may be is more appropriate model than the no-admixture model, or FLOCK, when the sample contains such individuals. that are admixed between populations."

Page 17, Line 318. "... with correlated allele frequencies seems like it would seem an interesting challenge, ..."

Page 22 List of Figures. Lines 386 - 393. Program DISTRUCT plots of cluster membership probabilities of steelhead individuals (q_i values from STRUCTURE and the normalized likelihood values from FLOCK). Individual fish are grouped left to right into their geographic samples, and each horizontal plot represents a run of the program

STRUCTURE, or the “best run” from a series of runs in FLOCK. Within horizontal plots, each steelhead individual is represented by a vertical bar composed of colored segments of lengths proportional to the various cluster membership probabilities. The numbers in parentheses ... same visually indistinguishable ‘best run.’23