ABSTRACT

Lines 30-31: Reword sentence; suggest “Use of SNPs has been favored due to their abundance in plant and animal genomes, ...”

Lines 31-32: Reword, “Here we present *in vitro* (obtained from targeted sequencing) and *in silico* discovery of SNPs and the design...”

Line 35: get rid of “thus” after “were”

Line 36: remove comma after “arrays”

Line 42: remove comma after “0.3”

Line 44: change “Such panels might make” to “Such panels might provide”

INTRODUCTION

Lines 49-50: There’s no need the sentence on SNPs in humans.

Line 54: replace “serve” with “are useful”

Lines 58-59: The transition between these two paragraphs is abrupt. The second paragraph in the introduction could benefit from a different topic sentence that makes the transition easier.

Lines 60-61: I think the percentage of molluscs in total aquaculture production is much higher than 27%. There’s also a new FAO report out for 2013 that you should cite instead of the 2009 version.

Line 65: Update FAO citation

Line 67: remove comma after “populations”

Lines 68-70: This sentence is confusing and needs rewording. I suggest “This work began with the use of allozymes and has since expanded to include mitochondrial sequences, AFLPs, and microsatellite markers. Several hundred markers are now available in *C. gigas* allowing for the development...”

Line 74: remove comma after “SNPs”

Line 78: Change “genome” to “genomes”

Lines 78-80: Reword this sentence or combine it with another sentence about oysters so that you are directly comparing the statistics with oysters. If you use “SSR”, please define it.

Line 86: change “issue” to “difference”

Line 95: add comma after citation and remove “and can reach”

Lines 96-97: remove from “depending on” through “SNPs”

Line 107: change “hundreds” to “thousands” if the example you give is 3072

Lines 107-108: This is the first time you mention SNP conversion rate and it would help the reader if you defined it

Line 113: remove comma after “flat oysters” and replace “have been obtained” with “already exist”

Line 114: insert comma after citations, remove “and allow” and change to “allowing for”

Line 177: insert comma after “structure”

Line 120: change “developing” to “being developed”

General comments: Considering how SNPs were chosen for analysis (by mining ESTs), is it worth mentioning that microsatellites are rarely associated with specific ESTs?

You cite a number of instances in which microsatellites are more powerful than SNPs, it might be good to include a sentence on why you would choose SNPs even if they are less informative markers. One way of doing this may be to address the unreliability of microsatellites, i.e. allelic dropout

MATERIAL AND METHODS

Line 126: change “consisted of” to “came from”

Line 128: change “resistance” to “resistant”

Line 129: change “less” to “fewer”

Line 138: change “Willington” to “Wilmington”

Line 142: change “several” to “a subset of”

Line 145: define “ESTs” for the first time

Lines 149 and 151: for numbers of ESTs make n = 61, n = 42

Line 152: Were all the ESTs sequenced as described in Sauvage et al.? The way the sentences are constructed it seems as if only the 42 were done this way. If the first 61 and the last 5 were done differently, please include that information.

Lines 155-156: Were these 5 additional SNPs added for all 24 oysters, or just the 10? This sentence should also be moved up to where you describe the other ESTs sequenced.

Lines 157-158: include the date (year) when you accessed the database

Line 160: what were the assembly parameters?

Lines 162-167: change “minimal” to “minimum”

Line 163: fix “at its the”; define what you mean by “position”

Lines 162-164: If you don’t include the results from this analysis in your manuscript I don’t see a need for these details in the methods.

Line 165: replace “allow us to get enough” with “produce many”

Line 167: replace “another SNP...analyzed one” with “a SNP within 120 bp of the SNP of interest”

Lines 172-173: rephrase to “to the first generations of three selected families” and delete text in parentheses

Lines 179-187: Why is this amount of detail included if the work was done in Cahais et al.?

Line 190: rephrase “we only used...(227)” to “we only used SNPs with the highest score of 227”

Lines 193-194: replace “to carry out the” with “for”

Line 197: was there only one negative control on each slide?

Line 206: when you write “sample” do you mean oyster? Please clarify

Lines 209-211: rephrase to samples were discarded if they showed...all the samples.”

Line 210: is the plus or minus 0.01 a percent?

Line 219: please define what the different MAF values mean, especially since you discuss them at length later on

Line 222: “In the Pacific oyster”

Line 227: delete “here,”

Line 228: Change “The method” to “This method”

Lines 230-231: Insert numbers in front of the 4 situations

Line 234: change “generated” to “simulated”

Line 235: are 1% and 5% typical values?

General comments: Especially since you mention sampling locations and methods later in the manuscript, it might be helpful here to mention how oyster tissues were samples. I would also like to see a map (as a supplementary figure) of where the wild populations are located.

It might help the reader to divide the “SNP discovery and selection for array construction” section into two parts – one for each species.

Make sure you define Illumina functionality score in the methods

In the “SNP genotyping array” section please state that all the oysters were genotyped on the arrays

RESULTS

Line 224: change “but” to “and”

Line 254: change “few” to “some”

Line 260: fix bracket in front of 0.9

Line 280: replace “independency” with “independence”

Line 296: change to “lower allele frequencies”

Lines 296-300: Please also present the MAF results for *O. edulis*

Lines 304-308: From “When all” to “200 SNPs” was already clearly stated in the methods. No need to restate here.

Line 308: Change “From 150 markers” to “With 150 markers”; delete “we can observe that”

Line 309: change “can be” to “was”

Line 310: delete “we can conclude that”

DISCUSSION

Line 326: make “SNPs” into “SNP”

Lines 326-328: This topic sentence does not accurately reflect the paragraph it precedes. Please reword. Also, this paragraph is very long and would be easier to read if it was reorganized.

Line 332: Change “enough SNPs” to “enough usable SNPs”

Lines 332-333: delete “to be submitted to Illumina”

Lines 333-335: Reword sentence, I suggest: “...functionality score (Fig. 2) where *C. gigas* had a lower proportion of *in silico* SNPs between 0.9-1 (60%) compared with the 100% of *in silico* SNPs with a score of > 0.9 in *O. edulis*”

Line 338: what does +30% mean? Change “bad” to “poor”. Delete “for those in vitro O. edulis SNPs”

Line 339: change “we observed” to “there was”

Line 340: What does 10% refer to – failure rate? Variation?

Line 340-341: Reword starting with “distribution” through end of sentence: “Illumina functionality scores were very similar between *in vitro* SNPs for *C. gigas* and *O. edulis*”

Lines 341-343: Explain why your data did not follow the expected pattern

Line 343: delete “be in” and “of”

Line 250: Why were the oysters not included in the genotyping array? Delete “Unfortunately”. Define which SNPs (instead of “those”).

Lines 351-352: remove phrase in parentheses

Lines 356-362: This entire paragraph can be moved to the methods. You should also include a sentence about the oyster genome.

Line 363: Delete “In addition”

Line 364: Change “of the” to “for”

Lines 364-366: remove score results because they are already stated in the results section

Lines 366-367: what are “those rates”?

Line 367: Define which clusters you are referring to

Line 369: delete “relatively”

Lines 369-372: rephrase the sentence to say something like “The oyster rates fall within the range of values seen in other non-model species...” because it’s a little confusing as written

Line 372: Change “These scores” to “The success rates”

Lines 373: change “the success rates” to “those”

Lines 374-376: no need to list each species, just list the citations

Line 376: Are you referring to conversion rates in model species?

Line 377: what do you mean by “this” rate?

Lines 378-379: are maritime and sugar pines non-models?

Line 380: what is “it” that reached 86%?

Line 382: In all other instances in this paragraph you use common names – please maintain consistency

Lines 372-382: please be specific about the rates to which you are referring in each sentence. A better organization may be to first discuss conversion rates and then make a clear switch to success rates.

Line 387: SNPs should be SNP

Lines 385-388: can you put this in the context of Figure 1?

Line 389: Delete “Finally” and replace “appears to be” with “is”

Line 392: Replace “appears that” with “even”

Line 395: Why is there a range of 30-50%? Does this encompass the two species? If so, please state that clearly.

Lines 397-400: Rephrase to “The relationship between Illumina score and genotyping success is in agreement with a number of other studies performed in non-model species and supports Illumina’s recommendations of using SNPs with functionality scores above 0.6.

Line 400: Put “the” between “in oyster”

Line 401: Change “seems to be” to “was”

Line 403: remove comma after “noted”

Lines 403-405: move to methods

Line 407: delete “such” and “in different kinds of studies”

Line 408: replace “our” with “the”

Lines 410-411: rephrase to “consequence of inclusion of *in vitro* SNPs”

Lines 411-412: Please explain this sentence because its meaning is unclear

Line 413: change “was of the” to “was on the”

Line 415: change “this” to “the”

Lines 415-416: italicize *in silico*

Line 417: change “discovery of” to “discovery between”. Change “It might” to “The contrast might”

Line 418: italicize *in silico*

Lines 410-417: This section needs to be clarified. Are you saying that most of the low MAF SNPs are from *in silico*? Lead with that if you are. Also, it gets a little confusing when you switch from comparing *in vitro* and *in silico* to comparing species. Please be very specific about what you are discussing.

Line 424: Specify what you mean by regional (European could be regional) and change “scale” to “scales”

Line 427: rephrase “characterization of its closely related sub species *C. angulata*”

Line 430: delete “these”

Line 432: rephrase to “QTLs associated with survival during summer”

Line 434: delete “already”

Line 435: replace “currently developing” with “in progress”. Do you mean phenotyping or genotyping?

Line 437: remove “that might be”

FIGURES

Figure 1 legend: is a sample an oyster? Please specify.

Figures 2 and 5: Make sure the brackets on x-axis are correct

Figures 2 and 3: I think these would be easier to visualize if each species was one color and then in silico and in vitro were represented by a different pattern.