SCRS/2020/068 Collect. Vol. Sci. Pap. ICCAT, 77(2): 312-324 (2020)

CALIBRATION OF THE FISH AGEING SERVICES READINGS CARRIED OUT IN GBYP PHASE 7, TO ESTIMATE AGE OF BLUEFIN TUNA FROM THE EASTERN ATLANTIC STOCK

E. Rodriguez-Marin[1], P. Addis[2], R. Allman[3], A. Bellodi[2], D. Busawon[4], F. Garibaldi[5], P. L. Luque[6], P. Quelle[1]

SUMMARY

The Fish Ageing Services laboratory (FAS) was contracted by ICCAT GBYP in Phase 7 to provide age estimates from 2000 Atlantic bluefin tuna otolith samples. With the objective of ensuring that age readings provided by FAS follow the ICCAT reviewed reading protocol, a sub-sample calibration exercise was carried out. The findings show that band counts are similar between FAS and the group of laboratories involved in direct ageing. This is reflected by an acceptable precision between both readings. However, there is a one-year bias in the count of bands in older specimens, starting from 10-13 years of age, with a lower count by FAS compared to the rest of the laboratories. This bias seems to be due to the fact that FAS counts the bands in a different area of the otolith ventral arm than other laboratories. This counting discrepancy, although small, is significant and it would be necessary for FAS to reread the samples of specimens older than 10 years using the area close to the sulcus margin of the ventral arm.

RÉSUMÉ

Dans le cadre de la phase 7 du GBYP, l'ICCAT a passé un contrat avec le laboratoire Fish Ageing Services (FAS) afin qu'il fournisse des estimations d'âge de 2.000 échantillons d'otolithes de thon rouge de l'Atlantique. Dans le but de garantir que les lectures d'âge fournis par FAS suivent le protocole de lecture révisé de l'ICCAT, un exercice de calibrage d'un sous-échantillon a été effectué. Les résultats montrent que le nombre de bandes est similaire entre FAS et le groupe de laboratoires participant à la lecture des otolithes. Cela reflète une précision acceptable entre les deux lectures. Cependant, il existe un biais d'un an dans le nombre de bandes des spécimens plus âgés, à partir de 10-13 ans, FAS présentant un nombre inférieur par rapport au reste des laboratoires. Ce biais semble être dû au fait que FAS compte les bandes dans une zone du bras ventral des otolithes différente de celle des autres laboratoires. Cet écart de comptage, bien que faible, est important et il faudrait que FAS relise les échantillons de spécimens âgés de plus de 10 ans en utilisant la zone proche de la marge du sillon du bras ventral.

RESUMEN

El laboratorio de Fish Ageing Services (FAS) fue contratado por ICCAT en la Fase 7 de GBYP para proporcionar estimaciones de edad de 2000 otolitos de atún rojo del Atlántico. Con el objetivo de garantizar que las lecturas de edad proporcionadas por FAS siguen el protocolo de lectura revisado de ICCAT, se realizó un ejercicio de calibración de una sub-muestra. Los resultados indican que el contaje de bandas es similar entre FAS y el grupo de laboratorios involucrados en la lectura de otolitos. Esto se refleja en una precisión aceptable entre las lecturas. Sin embargo, existe un sesgo de un año en el recuento de bandas en especímenes a partir de 10-13 años de edad, con un recuento más bajo por parte de FAS. Este sesgo parece deberse al hecho de que FAS cuenta las bandas en un

[1] Spanish Institute of Oceanography. Santander Oceanographic Centre, Santander (Spain). enrique.rmarin@ieo.es
[2] Department of Environmental Life and Science, University of Cagliari (Italy).
[3] National Marine Fisheries Service, Southeast Fisheries Science Center, Panama City (USA)
[4] St. Andrews Biological Station, St.Andrews (Canada).
[5] Dipartimento di Scienze della Terra, dell'Ambiente e della Vita, University of Genova (Italy).
[6] AZTI, Pasaia, Gipuzkoa (Spain).

área diferente del brazo ventral del otolito. Esta discrepancia, aunque pequeña, es significativa y sería necesario que FAS volviera a leer las muestras de ejemplares de más de 10 años utilizando el área cercana al margen del surco del brazo ventral.

KEYWORDS

Age estimation, otolith, standardization, Thunnus thynnus

1. Introduction

The ICCAT GBYP and national programs have invested a considerable effort in sampling calcified structures of Atlantic bluefin tuna (*Thunnus thynnus*, ABFT) and in age estimation from otoliths and the first radius of the first dorsal fin (spine). Direct ageing allows for good estimates of age composition of the catch enabling good cohort tracking of catch-at-age. This is why, since 2017, the “bluefin tuna Species Group” has been trying to use the direct ageing method to obtain the catch at age matrix used in the population assessment (Anon. 2017).

Methodology standardization of age interpretation from ABFT calcified structures, addressing intra and inter laboratory reader consistency, has been a priority within the GBYP (Busawon et al., 2020). With this purpose, an international workshop on ABFT direct ageing was held at the beginning of 2019 (Rodriguez-Marin et al., 2020). This recent GBYP workshop had the participation of most of researchers currently involved in direct ageing of ABFT. In addition, a scientist from the Fish Ageing Services (FAS), with extensive experience in the field of fish ageing, as well the GBYP coordinator, were also involved. The workshop reviewed the current protocols for otolith preparation and age reading criteria, which allows for standardized ageing methodology for future studies.

The Fish Ageing Services laboratory (FAS) was contracted by ICCAT GBYP in Phase 7 to provide age estimates from 2000 Atlantic bluefin tuna otolith samples. With the objective of ensuring that age readings provided by FAS follow the ICCAT reviewed reading protocol (Rodriguez-Marin et al., 2020), a sub-sample calibration exercise, funded by GBYP Phase 9, was carried out. An additional objective was to create a new reference collection from the aged samples.

2. Material and methods

Six research centers participated in this calibration, four from Europe (AZTI, University of Genoa, UNIGE; University of Cagliari, UNICA; and Instituto Español de Oceanografía, IEO), one from United States of America (National Marine Fisheries Service, NMFS, Panama City Laboratory) and another from Canada (St. Andrews Biological Station, SABS). All agers have experience in direct ageing of Atlantic bluefin tuna using otoliths and contribute with age-length data to the assessment of this species. The calibration exercise consisted of participants reading a sub-sample of 10% of the otoliths previously aged by FAS, to determine a measure of inter-laboratory precision. A total of 223 otoliths were used for this calibration, the sample selection attempted to be representative of the set of 2000 otoliths read by FAS, including the entire size range and covering as far as possible all the months of the year. ABFT used in this exercise come from eastern Atlantic and central and western Mediterranean Sea. The straight fork length of bluefin tunas analyzed ranged from 27 to 268 cm (**Figure 1**). Otolith section preparations were imaged and two sets, physical otolith sections and digital images, were used in the calibration exercise. This enabled the participation of all research centers, as images are easier to share, and allows, to a limited extent, the comparison between both sets of readings. Additionally, physical sections and digital images will be used to build a reference collection using the consensus age obtained from readings.

According to the reviewed protocol (Rodriguez-Marin et al., 2020), age estimates consisted in the counting of opaque bands using transmitted light. The otolith sections, were read twice. A third and final age reading was completed if the two age estimations differed by 2 or more years, to produce the final age estimate. The 3rd and final age estimate was done with knowledge of the first two readings. All readings were performed blind without knowledge of fish size or catch date. A reading form was provided and the following information was recorded for each sample: number of annual bands (opaque), ventral arm edge type (wide translucent, narrow translucent or opaque), edge confidence (1= no confident; 2= confident in completeness and not with the type and 3= confident), readability code (1= pattern present-no meaning, 2= pattern present-unsure with age estimate, 3= good pattern present-slightly unsure in some areas, 4= good pattern-confident with age estimate), reading date and notes with observations about the sample.

The readings from each reader and laboratory were compared with FAS readings. At the IEO laboratory the ageing was based on two readers, each of which read the whole data set once, using the most experienced reader for the third reading when ages were different by more than one year. At the UNICA University two readers made independent estimates of age, each reader reading the whole data set twice. Three modal readings have also been used: "Mode Experts" (Mode_E) for the readings of all laboratories including the readings of both physical otolith sections and digital images; "Mode Picture" (Mode_P) for laboratories that used digital images and "Mode Live" (Mode_L) for laboratories that used physical otolith sections. FAS used live readings and for the rest of the laboratories the reading mode is indicated at the end of the name (lab. name_L for live and lab. name_P for pictures). Precision was estimated through Coefficient of Variation (CV), Average Percent Error (APE), Evans-Hoenig and Bowker symmetry tests, age bias plots and age difference distributions between readers (Campana et al., 1995; McBride, 2015). FSA, R package version 0.8.20 (Ogle 2018) was used for the analysis.

3. Results and discussion.

Diagnosis of paired age agreement was obtained by readers and readers mode (**Table 1**). CV values were close to 10, which is the precision level required for production ageing. However, this value was slightly exceeded in 4 comparisons. The precision between FAS and modal values show that the set of experts have acceptable precision compared to FAS. The symmetry tests showed bias in four cases regardless of the type of method used, physical sections or digital images, highlighting the labs UNIGE and NMFS, which showed bias in both tests of symmetry. The mean readability and edge type confidence by each reader compared with FAS showed high values, above the average for each scale (**Table 1**). The percentage of coincidence with FAS readings in the type of edge assigned was also analyzed, obtaining acceptable results, since there are three possible options. The greatest coincidence between labs in the edge type was obtained with the IEO and even a better percentage of agreement was obtained with "Mode Experts" (**Table 1**).

Although tests of symmetry showed no bias between FAS and Modal readings (**Table 1**), age bias plots indicate a slight bias as age increases (13+), with FAS age estimates being lower by one year than "Mode Experts" and "Mode Live" (**Figure 2**). Furthermore, this bias was apparent at an earlier age (10+) with “Mode Picture”.

Reading analysis showed that the following laboratories IEO_L, UNICA_2_P, UNICA_1_P and SABS_P provided similar results to FAS, although these labs tended to count more annual band for older specimens (13+) (**Figure 3** and **Figure 4**). Similarly, AZTI and UNIGE also had higher annual band counts compared to FAS. On the other hand, NMFS showed an opposite trend with lower annual band counts compared to FAS.

To determine the possible reason for the lower number of opaque bands counted by FAS laboratory in older specimens (counting different structures, false annulus or edge type interpretation), a detailed inspection of FAS annotated images was performed. This bias seems to be due to the fact that this laboratory counts the bands in a different area of the otolith ventral arm than the rest of the laboratories. FAS lab counts on the external margin of the ventral arm, while the other labs do so in the area between the sulcus margin and the ventral groove. Furthermore, they used the dorsal arm to corroborate or even read age, even though the use of the ventral arm was recommended in the ABFT age validation study as readings from the dorsal arm resulted in age underestimation (Neilson and Campana, 2008). An erroneous count of false annual bands by FAS does not appear to occur, although in some cases the first annual band is counted further away with respect to the first inflection than the other readers. Possibly the lack of a reference scale in FAS readings affects in this regard. The interpretation of the edge of the otolith influences all readers, regardless of the laboratory of origin.

The annotated images of NMFS lab were also analyzed to try to discern the lowest band count compared to FAS and other labs. The difference from FAS is less than 0.5 bands at almost all ages. This seems to be because NMFS lab sometimes uses the external margin of the otolith ventral arm to count or corroborate the number of bands identified in the agreed reading area and located between the sulcus margin and the ventral groove. Furthermore, NMFS lab sometimes interprets the pattern of the first 6-8 years with fewer bands than other labs, including possible misidentification of the first band and the pattern of deposition of the first 5 annual bands. The identification of the first annual band is decisive in detecting the pattern of gradual decrease in the distance between the first annual bands (Rodriguez-Marin et al., 2019).

The transverse bands that indicate the years in the ventral arm of the otolith, cover the entire width of the arm until the age of 4 years. From these 4 years the bands begin to be not easily identifiable along the entire width of the arm, because their curvature and the discontinuity of their path make it difficult to match the band count of both sides of the arm. For this reason it is important to use the same reading area inside the ventral arm. FAS's use of the external margin produces systematically smaller annual band count in 1 year than ICCAT readers, starting from 10-13 years, and this discrepancy can increase in older specimens. This difference, although small, may be significant because, for example, with FAS counting procedure, the very abundant cohort of 2003 may appear as the one of 2004. From these results, it would be necessary for FAS laboratory to reread the samples of specimens older than 10 years, within the ICCAT GBYP Phase 7 contract, using the area close to the sulcus margin of the ventral arm (**Figure 5)**.

The participation of a scientist from FAS in the international GBYP workshop on ABFT direct aging conducted at the beginning of 2019, allows for the use of a common reading criterion for the readings of the next 2000 otoliths commissioned by GBYP in Phase 9 to this ageing laboratory. The reading of these otoliths from both GBYP phases, after an age quality control, will allow to have 4000 values of length at age to be considered for the next stock assessment.

Tile plots for edge type assignment and confidence showed some small similarity between laboratories (**Figure 6** and **Figure 7**). The tile plot of the readability code generally showed good confidence with the readings, although the readings based on pictures seem to reflect a better quality compared to the readings based on live samples (**Table 1** and **Figure 8**).

A potential source of bias could be the light type, since most labs, except FAS and NMFS, use reflected light and readers from these labs reported difficulties in reading due to light type change. However, this is likely to be a minor factor as ageing exercises have shown no significant light type effect on age interpretation (Rodriguez-Marin et al., 2014). What seems clear is that readers showed better agreement on edge type compared to other exercises using reflected light, since it allowed to reach a 53% of agreement in the type of marginal edge between "Mode Experts" and FAS. This being the reason argued by the participants of the 2019 workshop on ABFT direct aging to recommend the change of type of light in the age reading procedure (Rodriguez-Marin et al., 2020). Another factor that could have influenced the difficulty of reading the live samples is the absence of a reference scale, whereas the traditional method of reading has so far been based on images with reference scale.

To obtain the reference collection, the samples have been examined according to their readability code and whether it was necessary to carry out a 3rd reading. Samples that had at least three readers, including both live and images readings, with the worst readability code (1) or that would have required a third reading due to differences greater than 2 bands between the first and second reading, have been removed (4 samples).

4. Conclusions

The findings show that band counts are similar between FAS and the group of laboratories involved in direct ageing of Atlantic bluefin tuna otoliths. This is reflected by an acceptable precision between both readings. However, there is a one-year bias in the count of bands in older specimens, starting from 10-13 years of age, with a lower count by FAS compared to the rest of the laboratories. This bias seems to be due to the fact that FAS counts the bands in a different area of the ventral arm of the otolith than other laboratories. Counting is slightly different if done on either side of the arm of the otolith, therefore, it is important to use the same reading area in the ventral arm. This counting discrepancy, although small is significant and it would be necessary for FAS laboratory to reread the samples of specimens older than 10 years using the area close to the sulcus margin of the ventral arm. A reference scale helps identifying the first annual band and consequently the following first annual bands. Given these results, a new reference collection with consensus ages and following the ICCAT reviewed reading protocol will be created.

Acknowledgements

This work was carried out under the provision of the ICCAT Atlantic Wide Research Programme for Bluefin Tuna (GBYP) Phase 9. We also would like to thank the National Marine Fisheries Service, Southeast Fisheries Science Center, (USA) and St. Andrews Biological Station, (Canada), for their collaboration in this calibration exercise.

References

Anonymous. 2017. Report of the 2017 ICCAT bluefin stock assessment meeting (Madrid, July 2017). SCRS/2017/010.

Busawon, D.S., Addis, P., Allman, R., Bellodi, A., Garibaldi, F., Ishihara, T., Karakulak, S., Luque, P.L., Quelle P. Rodriguez-Marin, E. 2020. Evaluation of Atlantic bluefin tuna otolith ageing protocols. Collect. Vol. Sci. Pap. ICCAT, 76(2): 147-171.

Campana, S.E.; Annand, M.C.; Mcmillan, J.I. 1995. Graphical and statistical methods for determining the consistency of age determinations. Transactions of the American Fishery Society. 124:131-138.

McBride. R.S. 2015. Diagnosis of paired age agreement: a simulation of accuracy and precision effects. ICES J Mar Sci; doi:10.1093/icesjms/fsv047.

Neilson, J. D. and Campana S. E. 2008. A validated description of age and growth of western Atlantic bluefin tuna (*Thunnus thynnus*). Canadian Journal of Fisheries and Aquatic Science 65: 1523-1527.

Ogle, D.H. 2018. FSA: Fisheries Stock Analysis. R package version 0.8.20.

Rodriguez-Marin E., Di Natale A., Quelle P., Ruiz M., Allman R. , Bellodi A., Busawon D., Farley J., Garibaldi F., Ishihara T., Koob E., Lanteri L., Luque P.L., Marcone A., Megalofonou P., Milatou N., Pacicco A., Russo E., Sardenne F., Stagioni M., Tserpes G. and Vittori S. 2014. Report of the age calibration exchange within the Atlantic Wide Research Programme for bluefin tuna (GBYP). ICCAT Standing Committee on Research and Statistics Documents SCRS/2014/150.

Rodriguez-Marin E., Quelle P., Addis P., Alemany F., Bellodi A., Busawon D., Carnevali O., Cort J.L., Di Natale A., Farley J., Garibaldi F., Karakulak S., Krusic-Golub K., Luque P.L., and Ruiz M. 2020. Report of the ICCAT GBYP international workshop on Atlantic bluefin tuna growth. Collect. Vol. Sci. Pap. ICCAT, 76(2): 616-649

Rodriguez-Marin E., Quelle, P., Busawon, D., and Hanke, A. 2019. New protocol to avoid bias in otolith readings of Atlantic bluefin tuna juveniles. Collect. Vol. Sci. Pap. ICCAT, 75(6): 1301-1314.

Table 1. Diagnosis of paired age agreement for all data (n = 223). Precision indices: CV = Coefficient of Variation, APE = Average Percent Error, Evans-Hoenig and Bowker symmetry tests, symmetry bias (*, ** = significant differences in one or both symmetry tests, $p < 0.01$), mean readability score, mean edge type confidence and edge type agreement with FAS readings. Readers acronyms are explained in material and methods section.

Readers comparison	CV	APE	Evans-Hoenig	Bowker	Symmetry bias	Mean Readability	Mean Edge type Confidence	Edge type agreement with FAS (%)
FAS-Mode_E	6,3	4,5	0,3698	0,0114				53
FAS-Mode_L	8,2	5,8	0,2548	0,0239				45
FAS-Mode_P	8,2	5,8	0,0266	0,0938				49
Mode_P-Mode_L	10,3	7,3	0,4332	0,1694				57
FAS-IEO_L	8,2	5,8	0,2861	0,0283		2,8	2,4	42
FAS-AZTI_L	12,7	8,9	0,0130	0,0008	*	2,7	2,2	34
FAS-UNICA_2_P	9,2	6,5	0,0109	0,1030		3,1	2,7	35
FAS-UNICA_1_P	10,4	7,4	0,2512	0,0475		3,1	2,6	37
FAS-SABS_P	10,6	7,5	0,0655	0,0031	*	2,6	2,4	38
FAS-UNIGE_P	9,6	6,8	0,0004	0,0052	* *	3,1	2,7	35
FAS-NMFS_P	7,9	5,6	0,0000	0,0006	* *	2,9	1,8	36

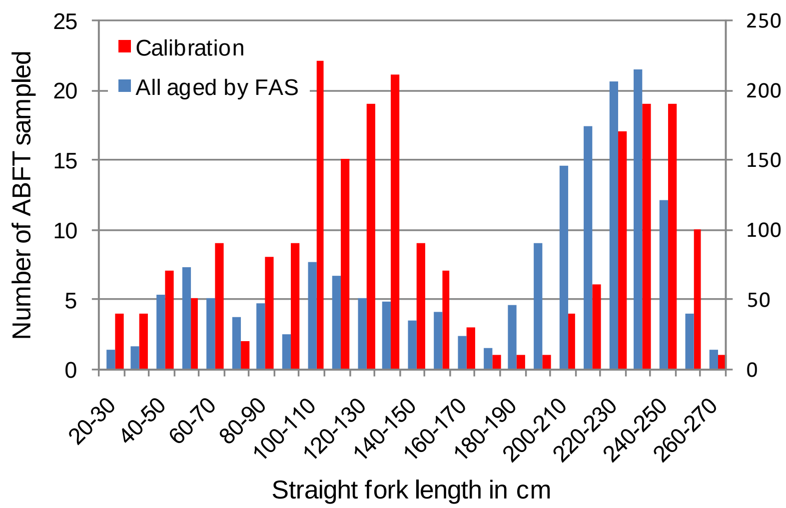

Figure 1. Length distribution of analyzed specimens by 10 cm size bin. The red columns represent the ABFT specimens used in the calibration exercise and the blue columns represent all the specimens whose otoliths were read by FAS.

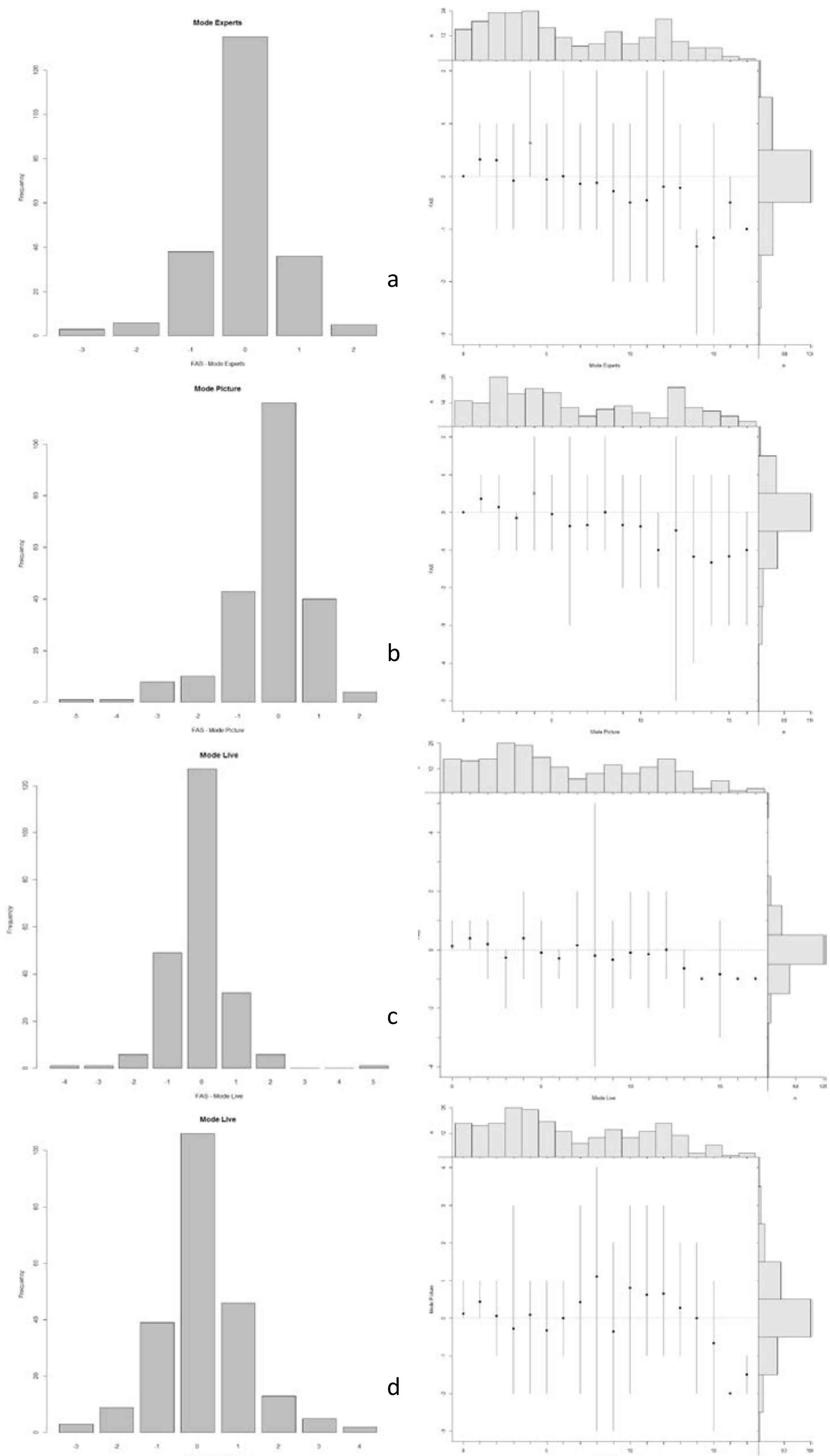

Figure 2. Age difference distributions and age bias plots between FAS and readers-Mode: a = "Mode Experts", b = "Mode Picture", c = "Mode Live", and d= between "Mode Live" and "Mode Picture". The number of samples per age class appears at the top of the bias plots.

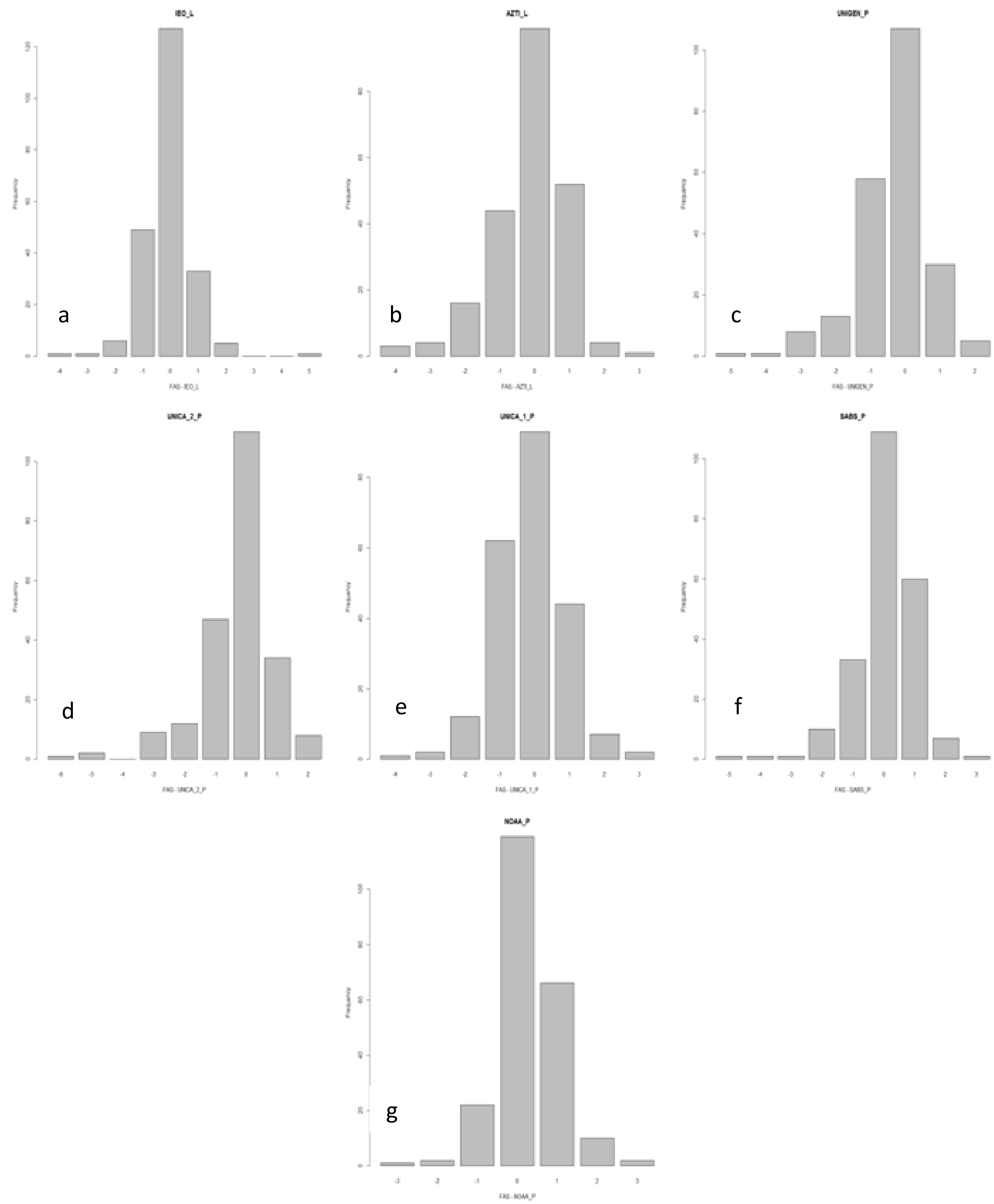

Figure 3. Age difference distributions between FAS and each reader-lab. a = IEO_L, b = AZTI_L, c = UNIGE_P, d = UNICA_2_P, e = UNICA_1_P, f = SABS_P and g = NMFS_P.

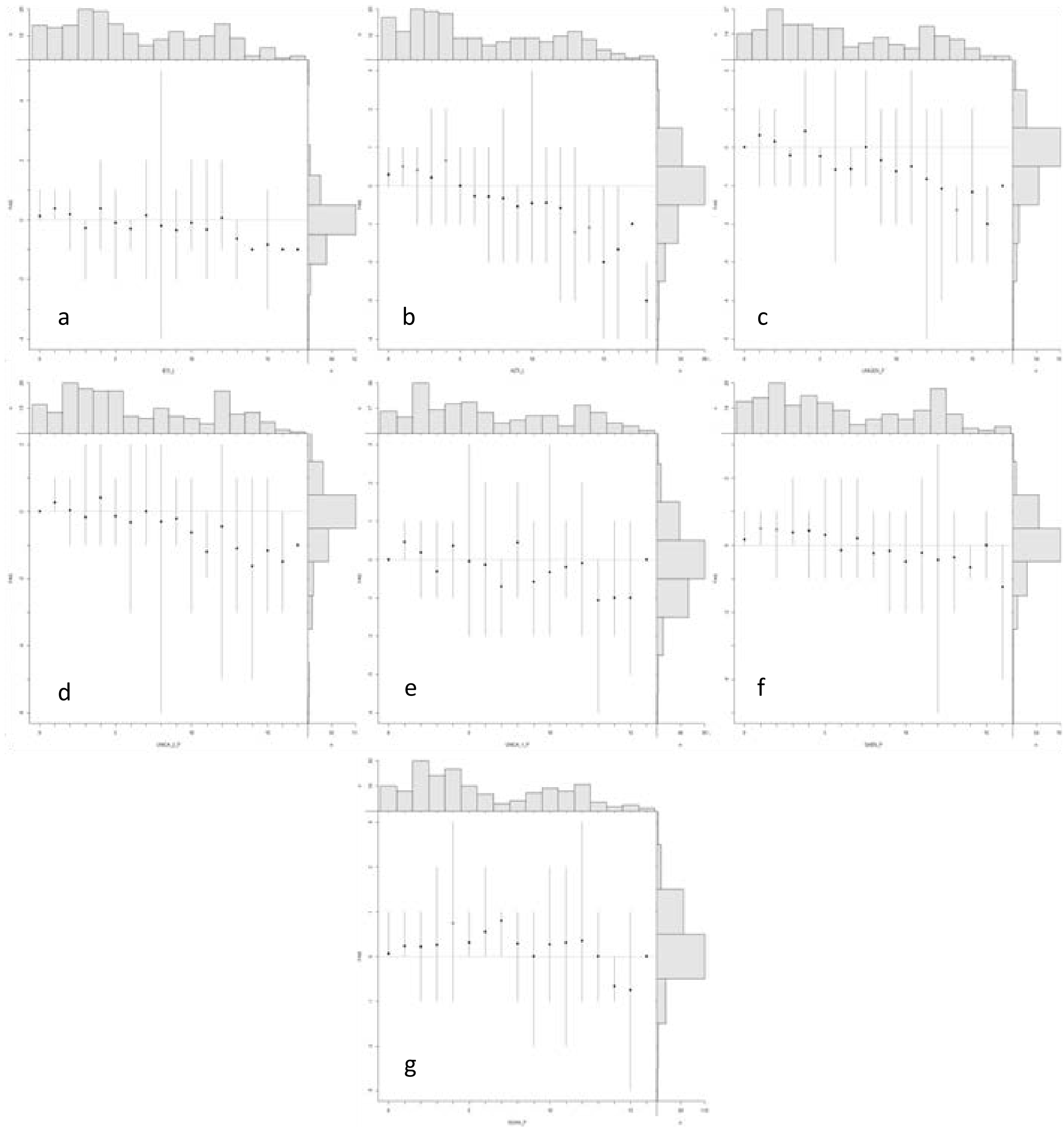

Figure 4. Age bias graphs (FAS age minus reader-lab. age). The number of samples per age class appears at the top of the graph. X-axis for figures: a = IEO_L, b = AZTI_L, c = UNIGE_P, d = UNICA_2_P, e = UNICA_1_P, f = SABS_P and g = NMFS_P. Y axis: FAS.

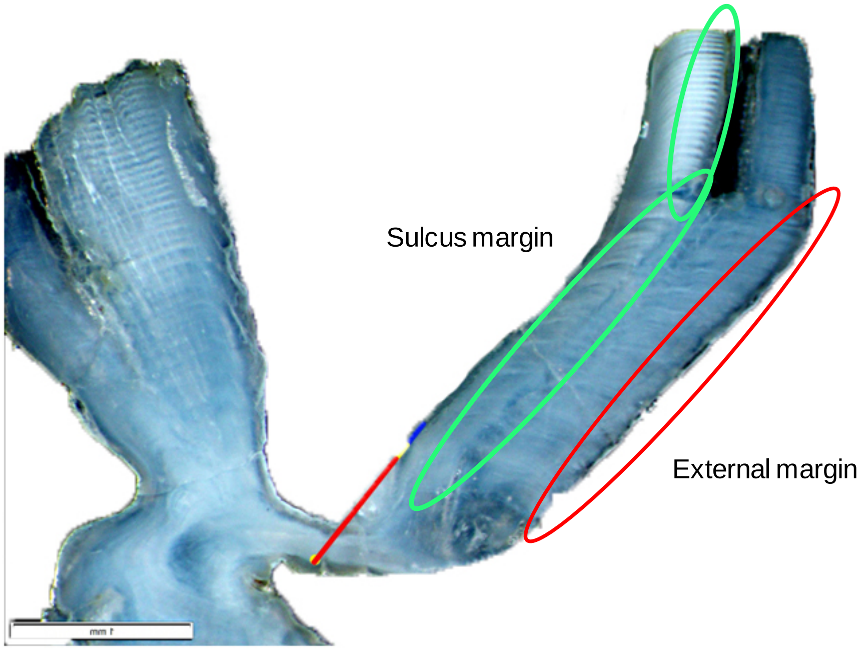

Figure 5. Otolith section with the identification of the margins of the ventral arm and the areas used to read the bands. In red the area used by FAS and in green the reading area used by the rest of the laboratories.

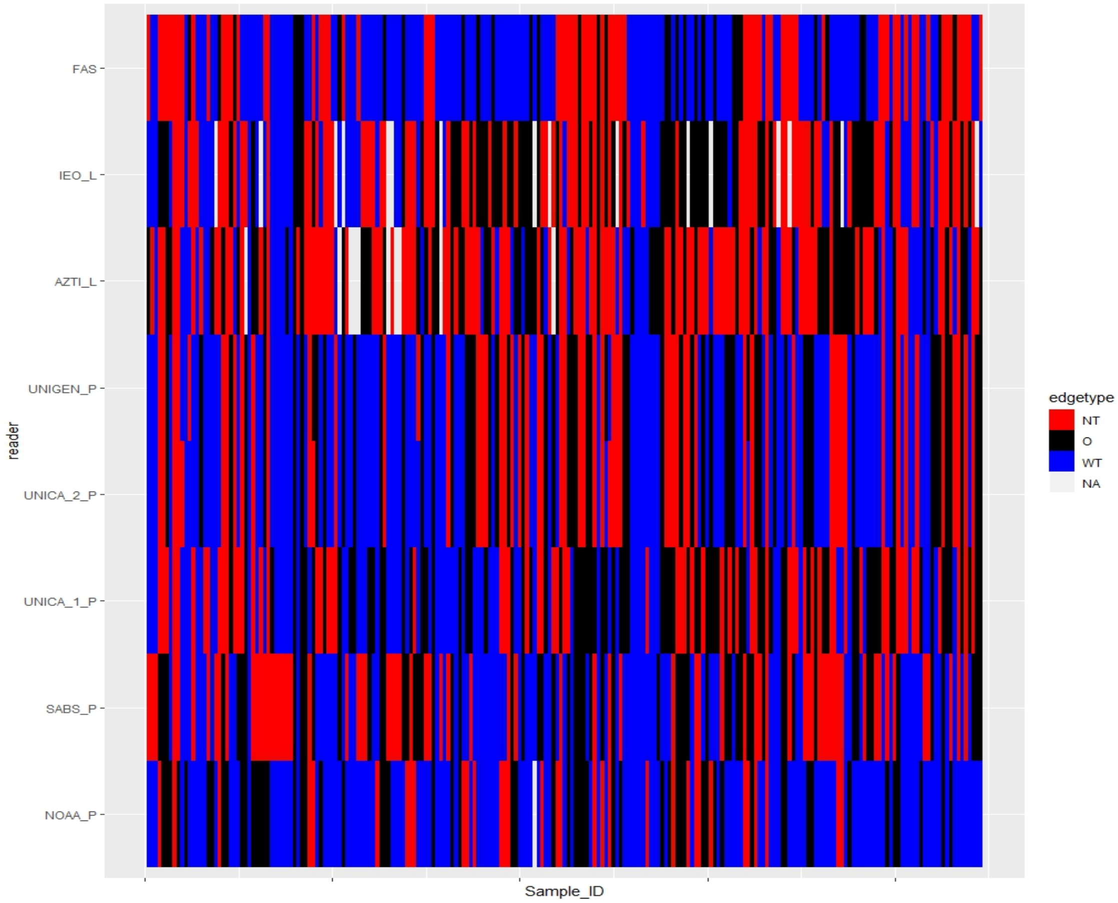

Figure 6. Tile plot showing otolith edge type assignment (NT= narrow translucent, Opaque= O, WT= wide translucent, NA= missing data) by sample for each reader-lab.

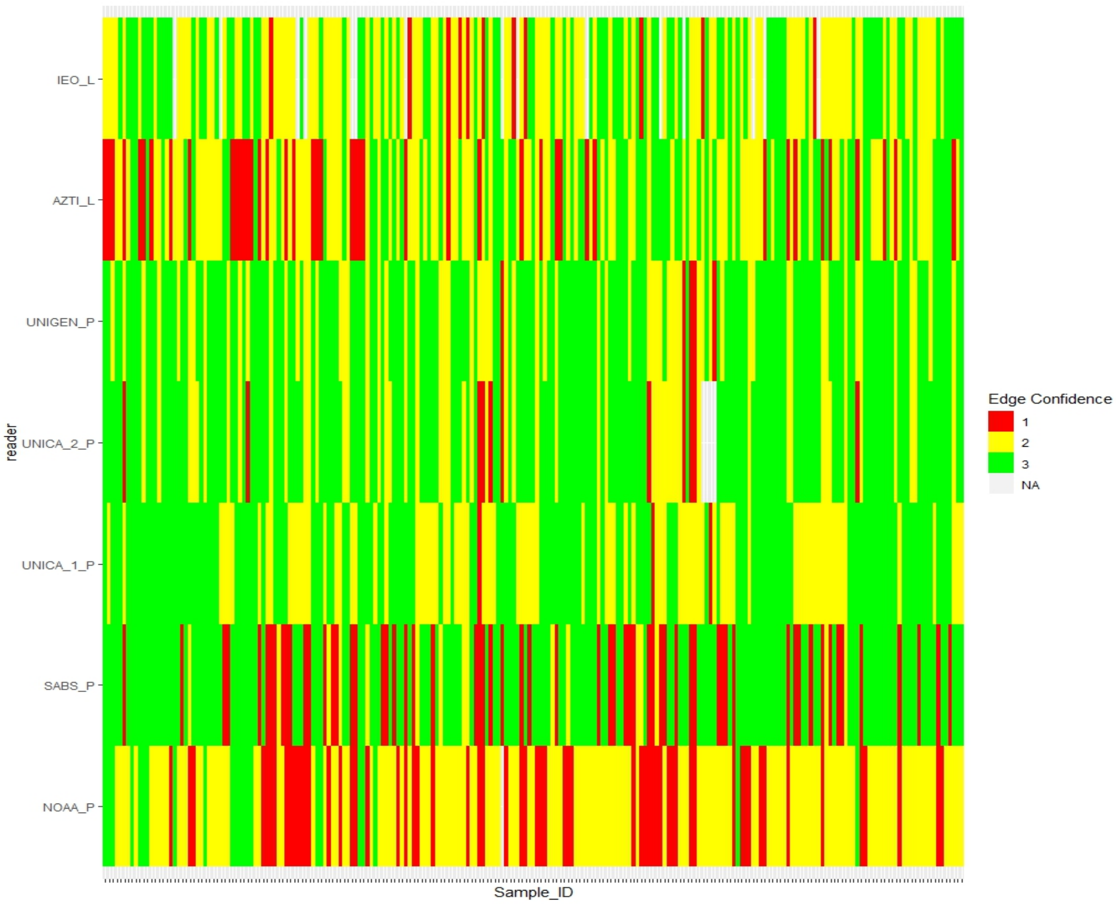

Figure 7. Tile plot showing otolith edge type confidence (1= no confident; 2= confident in completeness and not with the type and 3= confident) by sample for each reader-lab.

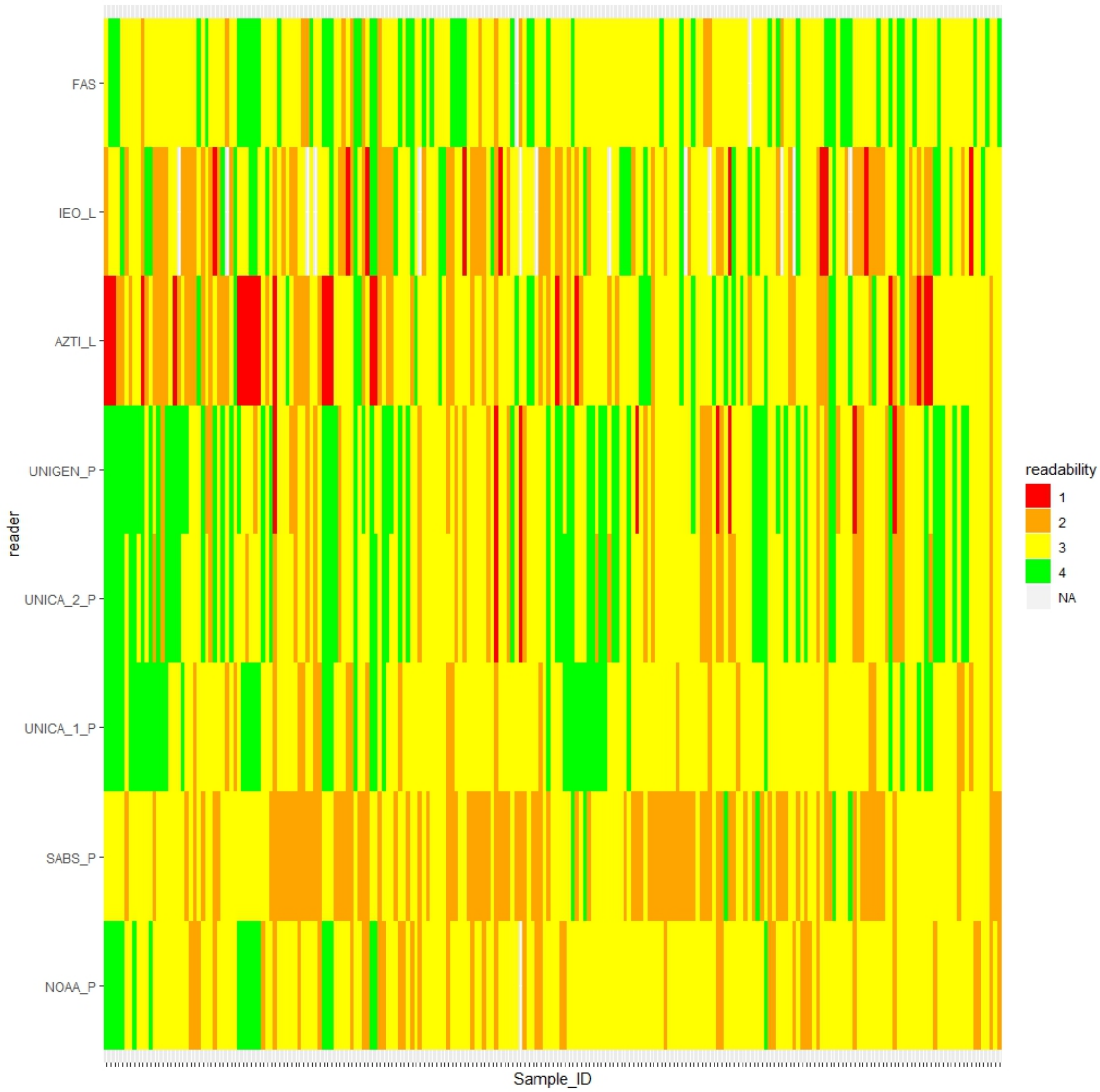

Figure 8. Tile plot showing otolith readability code (1= pattern present-no meaning, 2= pattern present-unsure with age estimate, 3= good pattern present-slightly unsure in some areas, 4= good pattern-confident with age estimate) by sample for each reader-lab.