

1 Inter-Laboratory Comparison of Fission Track Confined Length and Etch Figure Measurements  
2 in Apatite

3 Richard A. Ketcham<sup>1</sup>, Andrew Carter<sup>2</sup>, Anthony J. Hurford<sup>3</sup>

4 <sup>1</sup>Jackson School of Geosciences, The University of Texas at Austin, Austin, TX, USA

5 <sup>2</sup>Department of Earth and Planetary Sciences, Birkbeck College, Malet Street, London, WC1E  
6 7HX, UK.

7 <sup>3</sup>Department of Earth Sciences, University College, Gower Street, London, WC1E 6BT, UK,

8 Key words: GEOCHRONOLOGY: fission-track dating, track length measurement, apatite, inter-  
9 laboratory calibration, thermal history analysis

10 **Abstract**

11 Apatite fission-track length and etch figure data are powerful tools for obtaining thermal  
12 history information, but both require human analysts making manual measurements, and  
13 reproducibility is not assured. We report the results of an inter-laboratory study designed to  
14 clarify areas of congruence and divergence for these measurements and provide a basis for  
15 evaluating best practices to enhance intercompatibility of data sets. Four samples of megacrystic  
16 apatite from Durango, Mexico with induced tracks, one unannealed and three thermally  
17 annealed by varying amounts, were distributed internationally. In all, 55 analysts in 30  
18 laboratory groups participated in the experiment. Relative mean track lengths among the  
19 samples were consistent across all analysts, but measurements for each sample showed scatter  
20 among labs and analysts considerably in excess of statistical expectation. Normalizing  
21 measurements of annealed samples using the unannealed sample improved consistency, as did  
22 normalizing for track angle using c-axis projection. Etch figure data also showed variability

23 beyond statistical expectation, and consistency was improved by normalizing. Based on these  
24 data we recommend rigorous analyst training for length and etch figure measurement that  
25 includes measurement of standards, and that each analyst's data on unknowns be normalized by  
26 that analyst's own measurements on standards when using thermal history inverse modeling as  
27 part of the interpretation process.

## 28 **Introduction**

29 The key to resolving detailed thermal histories using the apatite fission-track (AFT)  
30 system comes from combining ages with track length data. Fission tracks form over time, and  
31 earlier-formed ones will experience more of a sample's thermal history than later-formed ones.  
32 This leads to characteristic patterns in horizontal confined track length distributions that can  
33 provide unique information on thermal history (Gleadow *et al.*, 1986). Even greater resolution is  
34 available when length data are paired with computational tools (e.g., Gallagher, 1995; Gallagher,  
35 2012; Green *et al.*, 1989; Ketcham, 2005) to identify the range of thermal histories that are  
36 consistent with both the length and age data and other geological constraints, using kinetic  
37 models of fission-track annealing (Crowley *et al.*, 1991; Ketcham *et al.*, 2007b; Ketcham *et al.*,  
38 1999; Laslett and Galbraith, 1996; Laslett *et al.*, 1987).

39 The ability to use track length data correctly and confidently hinges on the fidelity of the  
40 length measurements, and in particular their consistency with respect to the measurements  
41 underlying models of fission-track annealing (e.g., Barbarand *et al.*, 2003a; Carlson *et al.*, 1999;  
42 Green *et al.*, 1986). Although the analytical procedures used in these studies can be reproduced,  
43 and we understand many of the geometric sources of bias associated with track observation  
44 (Galbraith, 2005 Chapter 8; Galbraith *et al.*, 1990; Ketcham, 2003), full compatibility is not  
45 assured. In particular, because confined tracks are found and measured by a human analyst using

46 a microscope, rather than some mechanical or automated procedure, reproducibility of length  
47 data is an important concern.

48         The reproducibility of confined length data has been considered at both the inter-lab and  
49 intra-lab level. Intra-lab studies are valuable because they enable better control of conditions  
50 (such as repeating measurements of the same material using the same instrumentation), and thus  
51 allow focus on variables of interest. For example Green et al. (1986) include and compare  
52 measurements by two analysts of the same mounts. Barbarand et al. (2003b) gathered a series of  
53 induced-track data aimed at examining various aspects of reproducibility in detail, including  
54 among analysts and for single analysts over time, as well as the effects of Cf irradiation and the  
55 number of measurements necessary to converge to the correct mean length. Ketcham et al.  
56 (2009) obtained data for several analysts on two samples with induced and spontaneous tracks,  
57 and assessed the effects of measurement variability on thermal history reconstruction and the  
58 potential mediating ability of normalizing for length and angle.

59         Inter-laboratory experiments, however, provide the information necessary to assess the  
60 fidelity of measurements across the community and thus the overall reliability of the technique as  
61 it is applied. There has only been one previous large-scale inter-laboratory experiment for length  
62 measurements (Miller *et al.*, 1993), and its outcome was mixed, indicating general agreement but  
63 scatter substantially in excess of statistical expectation. Definitive interpretation was  
64 complicated, however, because the material was non-ideal, consisting of aliquots of natural  
65 samples with spontaneous fission tracks, which could conceivably have been non-homogeneous.  
66 Also, no information was reported about laboratory techniques, such as etching method.

67         With the help and cooperation of the international AFT community, we have performed a  
68 new inter-laboratory experiment designed to gather information intended to clarify areas of

69 congruence and divergence, provide a basis for evaluating best practices to enhance  
70 intercompatibility of data sets, and suggest areas that merit further study.

## 71 **Methodological Overview**

### 72 *Length revelation*

73 Fission tracks are too narrow (8-9 nm diameter in apatite; Li *et al.*, 2010) to be observed  
74 under an optical microscope, and instead are revealed by their ability to etch more easily than  
75 bulk crystal. Fission-track mounts are prepared by mounting grains in epoxy, polishing to reveal  
76 interior surfaces, and then etching in nitric acid (HNO<sub>3</sub>) at some prescribed conditions of  
77 strength, temperature and duration. Apatite etches anisotropically (Green and Durrani, 1977),  
78 with faster etching in the direction of the crystallographic *c* axis, and in general stronger etchants  
79 (higher concentration and/or temperature) are expected to increase this anisotropy compared to  
80 weaker ones.

81 Fission tracks intersecting the polished surface only contain partial length information, as  
82 one of their ends is missing. Confined tracks are revealed when etchant travels down a pathway  
83 from the exposed surface and intersects a track in which both ends are within the solid crystal.  
84 When the etchant pathway is another fission track, the confined track is referred to as a TINT  
85 (track-in-track; Lal *et al.*, 1969), while if the pathway is a fracture or cleavage it is termed a  
86 TINCLE (track-in-cleavage, Bhandari *et al.*, 1971).

### 87 *Length measurement*

88 Confined track lengths are measured under a microscope at high magnification, generally  
89 1000x-1600x, using either a drawing tube and digitizing tablet or a camera and specialized  
90 software to measure the distance between track ends. As surface etch rates vary according to



91 crystallographic orientation it is normal practice to control for etching efficiency and measure  
92 tracks only on grains in which the crystallographic *c* axis is in the polished plane, as these have  
93 the lowest bulk etch rate. *C*-parallel sections can be determined by their aligned etch figures  
94 (Donelick *et al.*, 2005), and track orientation with respect to the *c* axis can be determined as the  
95 angle between the track and the etch figure elongation direction. Confined tracks must be close  
96 to horizontal, although limited inclination has little effect; the projection of a track dipping 10°  
97 with respect to the polished surface is within a factor of  $\cos(10^\circ)$ , or 0.985, of the true length. It  
98 is also possible to measure steeper tracks if the measurement system records the 3D location of  
99 the endpoints and accounts for the apatite refractive index, allowing a correction to be made,  
100 although one test has found non-horizontal tracks to slightly increase the standard deviation of  
101 the length distribution (Jonckheere and Ratschbacher, 2010).

102         For a fission track to be considered measurable, it must be completely etched, with  
103 clearly defined ends. The revelation of a track will depend on its angle relative to the *c* axis  
104 (hereafter denoted as  $\phi$ ) and the etching protocol; examples are shown in Figure 1. When a  
105 strong etchant (5.0 or 5.5 M) is used, tracks are most easily observed, and more likely to be fully  
106 etched, when they are in the  $\phi$  range of approximately 30-85° (Fig. 1A,B) (Barbarand *et al.*,  
107 2003b; Donelick *et al.*, 1999). Tracks at lower angles (Fig. 1C, D track 1; 1E track 1) are thin for  
108 their entire extent. Tracks close to perpendicular to the *c* axis ( $\phi > 85^\circ$ ) can look wide but slightly  
109 distorted (Fig. 1E, track 2) or initially wide near the intersection with the etchant pathway and  
110 pinched toward the ends (Fig. 1F). The distorted shape of track 2 in Fig. 1E may result in a  
111 slight error in angle determination when the line connecting the endpoints is not parallel to the  
112 length. In the case of thin or pinched tracks, the ends can be indistinct, making it difficult to tell  
113 whether they are fully etched or not; under-etched tracks should not be measured because they

114 may not record the full length, and instead could give a spurious signal of track shortening.  
115 These etching effects, combined with the anisotropy of TINT etchant pathways, impart a  
116 substantial bias upon which angular populations are measured (Galbraith et al., 1990; Ketcham,  
117 2003).

118         The anisotropy in etching diminishes for a weaker etchant (e.g. <2N HNO<sub>3</sub>), reducing  
119 somewhat the dissimilarity of confined tracks at different  $\phi$  angles. However, weaker etchants  
120 require longer etching times (>40s vs. 20s), and etching duration brings up additional  
121 considerations (Jonckheere et al., 2007; O'Sullivan et al., 2004). There is inevitably some time  
122 delay before confined tracks start to etch, which varies depending on the strength of the etchant,  
123 extent of subsurface penetration along a pathway (TINT or TINCLE) and solubility  
124 (composition) of the apatite. As a result, even tracks at favorable angles can be under-etched,  
125 and the variability in etching time contributes to variation in measured track lengths. For similar  
126 reasons, short tracks are somewhat more likely to be fully etched than otherwise equivalent  
127 longer tracks, partially counter-acting the biasing effect of longer lengths being more likely to be  
128 intersected and etched than shorter ones (Laslett *et al.*, 1982).

129         Other situations that can lead to erroneous measurements are shallow tracks that intersect  
130 the polished surface (Fig. 1C,D track 2), tracks with fluid-filled and thus obscured ends (Fig.  
131 1G,H) and opposite-dipping semi-track pairs that appear superficially to be single tracks. Of  
132 these cases, the first two will lead to erroneously short measurements, and the third can be either  
133 short or long.

134         The analyst measuring fission tracks must thus constantly keep in mind, and adhere to,  
135 strict criteria for determining which confined tracks should be measured and which should be  
136 bypassed. There is likely to be some variability in these criteria between analysts and lab groups,

137 which will contribute to variation. Finally, measurement by microscopy or via stored digital  
138 images requires some form of calibration. Methods employed include calibrated microscope  
139 scale bars (typically at 1-2  $\mu\text{m}$  resolution) or SEM diffraction gratings (typically < 0.4-0.6  $\mu\text{m}$   
140 resolution). Calibration should be made only for the area in which the features to be measured  
141 are placed. Typically this is in the center of the field of view in order to avoid the defocused  
142 peripheral regions. Systems that use digitizing tablets are also vulnerable to models that have an  
143 uneven spacing of grid wires. Another potential issue when a drawing tube is employed is the  
144 size of the LED spot used to demark track ends; most analysts center the spot over the ends, but  
145 some use an edge of the spot to attempt to make the measurement more precise. All of these  
146 aspects can contribute to systematic differences between laboratories.

#### 147 *C-axis projection*

148 In addition to anisotropy of etching, apatite fission tracks also show annealing anisotropy,  
149 with tracks parallel to the **c** axis annealing more slowly than tracks oriented along the **a** axis  
150 (Donelick, 1991; Donelick et al., 1999; Green and Durrani, 1977). At low to medium amounts  
151 of annealing, annealing rates vary smoothly between these two orientations, and length  
152 distributions are well-represented as ellipses on a polar plot (Donelick, 1991), although in detail  
153 the distribution may be slightly non-elliptical (Jonckheere et al., 2007). At high levels of  
154 annealing, tracks at high  $\phi$  angles begin to shorten much more quickly and disappear while low- $\phi$   
155 tracks persist, a process termed accelerated length reduction (Donelick et al., 1999).

156 As annealing progresses, annealing anisotropy leads to greatly increased dispersion in  
157 lengths of tracks that have experienced identical amounts of heating, as shown in experiments in  
158 which induced tracks are annealed (e.g., Green et al., 1986). To compensate for this, Donelick et  
159 al. (1999) introduced **c**-axis projection, which was subsequently refined by Ketcham (2003) and

160 Ketcham et al. (2007a). **C**-axis projection is a transform that converts each  $(l, \phi)$  measurement  
161 into an estimate of what the length of a track oriented along the **c** axis that had experienced the  
162 same annealing conditions would be,  $l_c$ , and the uncertainty in that estimate,  $\sigma_{l_c}$ . It can be more  
163 generally viewed as a means of removing the dispersion caused by annealing anisotropy,  
164 resulting in a more precise index of thermal input for a given track than length alone.

165         Creating a **c**-axis projection model consists of fitting ellipse radii,  $l_{c,fit}$  and  $l_{a,fit}$   
166 corresponding to the **c**-axis and **a**-axis directions, to sets of tracks measured in a series of  
167 experimental annealing runs. These data are then used to fit a four-parameter projection  
168 transform (Ketcham et al., 2007a). This process generally requires dozens of experiments to  
169 document all stages of annealing and overcome dispersion in the ellipse fits, and thus has only  
170 been done on the two largest experimental data sets, by Carlson et al. (1999) and Barbarand et al.  
171 (2003a). These studies used slightly different etchant strengths, respectively 5.5 M and 5.0M  
172 HNO<sub>3</sub>, and the difference between their respective **c**-axis projection models was attributed to  
173 etchant (Ketcham et al., 2007a). The present study tests whether this assertion is correct, or  
174 some other factor may be responsible for variation in observed anisotropy effects.

#### 175 *Etch figures*

176         Etch figures, the intersections of etched tracks with the polished surface (Fig. 1D), are  
177 useful for determining crystallographic orientation, as a proxy for inferring the effective  
178 annealing kinetics (Burtner *et al.*, 1994; Ketcham *et al.*, 1999) and estimating initial  
179 (unannealed) track length (Carlson *et al.*, 1999). The principal measured parameter is the  
180 diameter of the track parallel to the apatite **c** axis when it is in the polished plane ( $D_{par}$ ); the **c**-  
181 axis perpendicular diameter,  $D_{per}$ , is also of potential use, but is more difficult to measure  
182 reliably. The Ketcham et al. (2009) experiment found reproducibility among multiple analysts

183 measuring  $D_{par}$  on the same grain mounts to be poor. Sobel and Seward (2010) studied the  
184 problem under better conditions and in considerably more detail, and suggest protocols for  
185 executing and normalizing  $D_{par}$  measurements.

## 186 **Methods**

187 Preparations for the experiment began in 2004. Mark Cloos (University of Texas at  
188 Austin) provided a selection of lime-colored apatite crystals from Durango, Mexico, and three  
189 were selected as containing minimal defects. Each was heated at 500°C for 24 hours to totally  
190 anneal spontaneous tracks. Aliquots of each crystal were polished, etched and inspected to  
191 confirm total spontaneous track removal. Each crystal was sliced into ~1 mm plates parallel to  
192 the *c*-axis using a fine diamond saw. The sliced crystals were wrapped in aluminum foil and  
193 irradiated at the Lucas Heights reactor (Australia) in April 2004 using nominal thermal neutron  
194 fluences of  $2 \times 10^{16} \text{ ncm}^{-2}$ . Each crystal was irradiated in a separate reactor run (TE68, 70 and  
195 77) to keep the total mass of active material within acceptable limits. Each irradiation was  
196 monitored by inclusion of a CN-5 dosimeter glass with a mica detector. Substantial radioactivity  
197 was allowed to decay until early 2008.

198 In 2008, induced tracks in the samples were partially annealed and aliquots distributed to  
199 participating laboratories. Apatite from irradiation TE70 was designated DUR-1, apatite from  
200 irradiation TE77 was divided into two aliquots designated DUR-2 and DUR-3, and apatite from  
201 irradiation TE68 was designated DUR-4. Appropriate annealing conditions were estimated  
202 based on Barbarand et al. (2003a), and the furnaces, annealing rig, thermocouples and all  
203 procedures were identical to those used by Barbarand et al. (2003a).

204 For sample DUR-1 annealing conditions of  $288 \pm 2^\circ\text{C}$  and 10 hr (plus ~3 min equilibration  
205 time on loading) were chosen to produce a track-length distribution similar to exhumed

206 basement, with a mean length ( $l_m$ ) of 11.5-12  $\mu\text{m}$ . Measurements of a test aliquot gave an  $l_m$  of  
207 12.05 $\pm$ 0.07  $\mu\text{m}$  (n=100). Sample DUR-2 was left unannealed as a control sample containing  
208 full-length induced tracks. For sample DUR-3 the aim was to produce a broad track-length  
209 distribution as might be found in a subsurface sample, more complex to measure but with a  
210 statistically-adequate number of tracks. It was annealed at 310 $\pm$ 2 $^\circ\text{C}$  for 10 hr (equilibration on  
211 loading took about 3.5 min), and an initial test aliquot gave an  $l_m$  of 10.20 $\pm$ 1.10  $\mu\text{m}$  (n=101).  
212 DUR-4 was intended to simulate a volcanic-cooling type track-length distribution. It was  
213 annealed at 240 $\pm$ 2 $^\circ\text{C}$  for 10 hr (equilibration on loading took about 2 min), and an initial test  
214 aliquot gave an  $l_m$  of 13.80 $\pm$ 0.80  $\mu\text{m}$  (n=102).

215 Part of the intention behind using single apatite crystals was to attempt to overcome  
216 compositional variation between different grains of a large apatite concentrate. However, sample  
217 compositional heterogeneity within a single crystal could still produce variation between results.  
218 To help monitor such variation, the cut plates of each apatite were broken and the relative  
219 positions of each sub-piece recorded by simple letter-number co-ordinates (e.g. A1, C5 etc).  
220 Where possible each lab was sent the aliquots for each apatite (DUR 1,2,3&4) taken from the  
221 same relative position (A1, C5 etc.); however, this system broke down in later distributions when  
222 inadequate orientated material remained.

223 Fission-track laboratories known to the authors were contacted and aliquots provided to  
224 those who agreed to participate. All preparation of grain mounts and measurement was done at  
225 each participating laboratory using its standard operating procedures and instrumentation. A  
226 survey (Supplementary Data) was also prepared to accompany all aliquots, so that laboratories  
227 could report pertinent information such as etching procedures, measurement systems, and analyst  
228 experience as of the time of the measurement. Survey answers are summarized in Table 1.

229 Results were returned by email. Each data set by a single analyst for a single sample was  
230 given a code (1-4)-L(1-47)-A(1-6)[-Q(1-3)], where the first number indicates the DUR sample  
231 number, the second is the laboratory group, the third is the analyst at that laboratory, and the “Q”  
232 designation is used as needed when the analyst measured the same sample multiple times  
233 (example: 1-L13-A2-Q1 refers to sample DUR-1 analyzed by lab 13, analyst 2, first  
234 measurement).

235 Summary statistics were calculated for all data submitted. The angular distribution of  
236 tracks in each measurement for which angle data were reported was analyzed by fitting polar-  
237 plot ellipses to provide the intercepts with crystallographic **c** and **a** axes ( $l_{c,fit}$ ,  $l_{a,fit}$ ), following the  
238 method of Donelick (1991) and Donelick et al. (1999). In results for DUR-3 that showed  
239 evidence of accelerated length reduction, the shortened tracks, generally those  $<7-7.5 \mu\text{m}$  at high  
240  $\phi$  angles, were removed for ellipse fitting; numbers ranged from 0 to 16 deleted measurements,  
241 the average was 2 per experiment (or 1.7% of tracks measured), and the median was 1.

242 We also calculated  $l_{c,mod}$ , the mean of individually **c**-axis-projected lengths, using the two  
243 models given in Ketcham et al. (2007a), which characterize the Carlson et al. (1999) and  
244 Barbarand et al. (2003a) data sets; for brevity, these are respectively referred to as the C99 and  
245 B03 projection models in the discussion below. Which of these two models more closely  
246 represents the tendencies of a given analyst may be evaluated by the extent to which  $l_{c,fit}$  and  
247  $l_{c,mod}$  match, or in how they co-evolve with increasing annealing.

## 248 **Results**

249 In all, 30 laboratory groups participated in the experiment, with 55 analysts in total, 53 of  
250 which provided analyses of all four samples.

251 To evaluate variability among aliquots, two analysts in lab 32 measured three separate  
252 aliquots. Lab 32 also included a virtually untrained analyst (number 4), who received only  
253 enough instruction to recognize a track and measure it, as an intentional end-member case of  
254 minimal experience.

255 Some laboratory groups independently decided to conduct additional measurements to  
256 capture additional information. Lab 13 re-polished and re-etched each mount three times with  
257 different etching protocols to inspect etching effects. Three analysts in lab 14 measured both  
258 TINT and TINCLE tracks and provided summaries for each; all results are reported, but for  
259 comparing among lab groups we utilize the combined results. Lab 41 performed measurements  
260 with and without  $^{252}\text{Cf}$  irradiation (Donelick and Miller, 1990) to enhance detection of confined  
261 tracks.

## 262 *Survey*

263 Responses to the survey are provided in Table 1. We requested for one survey to be  
264 submitted per analyst, although some lab groups submitted combined surveys, and others omitted  
265 some questions or neglected to submit them. Experience of users represents a full continuum  
266 from 38 years of experience to novice. Among those who reported at least 1 year of experience,  
267 the mean was 12.5 years.

268 Results indicate a number of areas of congruence in the community. Only one lab uses  
269 oil-immersion microscopy, and almost all used the straightforward method of demarking track  
270 ends directly with an LED or pointer when measuring, as opposed to previously mentioned  
271 strategies aimed at compensating for non-negligible LED size. Most analysts measured only  
272 TINT tracks to avoid uncertainties associated with fracture movement that could increase the  
273 apparent length of tracks and to avoid additional orientation bias (since cleavage in apatite is



274 oriented at {0001} and {1010}), as well as the possibility that geologic fluids could have  
275 infiltrated and pre-etched or otherwise fixed tracks at some earlier stage in their history  
276 (Jonckheere and Wagner, 2000).

277 Results showed an unexpected diversity of etching protocols, however. In all 14 different  
278 protocols were reported, although two of these were experiments intended to test etching effects,  
279 and one was an adapted procedure intentionally analogous to that used for zircon (e.g., Yamada  
280 et al., 1995). Seven employed 5.0 or 5.5 M HNO<sub>3</sub> at slightly different temperatures; seven used  
281 other etchant strengths. The number of track lengths measured on each sample varied among labs  
282 and samples, from 50 to over 200.

### 283 *Unannealed sample, mean length*

284 There was considerable variation among results for the unannealed sample, DUR-2. Data  
285 are given in Table 2, and mean track lengths and errors are plotted against several variables in  
286 Figure 2. Excluding four outliers below 15  $\mu\text{m}$ , there was a spread in results of almost 1.6  $\mu\text{m}$   
287 (15.25-16.84  $\mu\text{m}$ ). Overall, only 23 of the 65 measurements reported (35%) are within 2  
288 standard errors of the overall mean (15.89 $\pm$ 0.12  $\mu\text{m}$ ).

289 Experience and frequency of making measurements do not seem to be factors  
290 contributing to consistency of results; the most-frequent analysts (lab 13) differed from the  
291 second-most frequent (lab 14) by well over 1  $\mu\text{m}$  (Fig. 2B). There is a vague trend of increasing  
292 mean length versus operators' frequency of analyses and years of experience (Fig. 2B, C), but  
293 overall the data are scattered in this regard.

294 Interestingly, there is no evidence of a trend in initial length versus etching protocol, and  
295 in particular etchant strength (Fig. 2D). Overall the amount of variation observed using the two

296 primary etching protocols employed currently (5.0 or 5.5 M HNO<sub>3</sub>, 20s, 21°C) spans the range  
297 observed with different etchant strengths and temperatures and durations.

#### 298 *Annealed samples, mean length*

299 Figure 3 shows the mean lengths and uncertainties reported for the annealed samples, and  
300 data are provided in Tables 3-5. The degree of scatter is similar to or somewhat worse than  
301 obtained for the unannealed sample: in order of increasing degree of annealing (DUR-4, DUR-1,  
302 DUR-3), 16 of 62 (26%), 19 of 62 (31%), and 24 of 63 (38%) reported means were within two  
303 standard errors of the respective overall means.

#### 304 *Angular data*

305 Of 55 analysts, 42 reported angular data and 13 did not. The ellipse fits to each  
306 experiment are shown in Supplemental Data Figures S1-S5. For each sample, a characteristic  
307 pattern of length versus angle is observed across many laboratories and analysts. In general,  
308 results repeated patterns observed in the data of Carlson et al. (1999) and Barbarand et al.  
309 (2003a): generally an elliptical distribution, with the possible exception of the most-annealed  
310 sample DUR-3.

311 Two selected sets of ellipse fits for the four samples are shown in Figure 4; Fig. 4A-D  
312 show the results for a very experienced analyst from lab 32, and Fig. 4E-H show the  
313 corresponding fits for the novice user from the same lab. The measurements by the experienced  
314 analyst cluster tightly around the ellipses, while the measurements by the novice are much more  
315 scattered. Most of the novice's outliers were short compared to the experienced analyst's data,  
316 although some were long. The analysts also diverge in that there is little systematic anisotropy in  
317 the measurements of the novice, whereas the experienced analyst shows the familiar pattern of

318 increasing anisotropy at increasing levels of annealing (e.g., Donelick, 1991; Donelick *et al.*,  
319 1999).

320 The most-annealed sample, DUR-3, experienced conditions that put it into the first stages  
321 of the “accelerated length reduction” regime (Donelick *et al.*, 1999; Ketcham *et al.*, 1999), in  
322 which tracks at high angles to the **c**-axis begin to anneal more quickly, departing from the  
323 elliptical trend. Whether these shortened tracks were measured or not was an area of particular  
324 divergence. Figure 5 shows six additional examples (to go with Fig. 4C and G), in which all  
325 analysts had at least 19 years of experiences. These analysts range from measuring zero (top  
326 row) to a few (middle row) to several (bottom row) tracks that do not fall on the elliptical trend.  
327 In some cases (e.g., labs 41, 47), different analysts observing the same mount measured very  
328 different proportions of non-elliptical tracks.

329 The data also show divergence in the relative frequencies of measurement of elliptical-  
330 trend high- and low- $\phi$  tracks. For example, some analysts measured very few high-angle ( $\phi > 85^\circ$ )  
331 tracks (Lab 5-A1; see Fig. S1 for examples in this paragraph), and others measured very few in  
332 the unannealed samples but more in the annealed samples (Lab 5-A2, 34, 41). Some measured  
333 very few to zero low-angle tracks ( $\phi < 30-40^\circ$ ) in all experiments (Lab 20, 22, 30, 32-A1, 34, 41),  
334 while others measured an increasing proportion of low-angle tracks as annealing progressed (Lab  
335 7, 13, 14, 26, 32-A2), and others measured them with roughly equivalent frequency in all  
336 samples (Lab 5, 28).

### 337 *C*-axis projected data

338 Both **c**-axis projection models provided  $l_{c,mod}$  values that matched the fitted ellipses fairly  
339 well. The B03 model fits best, with a mean residual ( $l_{c,fit} - l_{c,mod}$ ) of  $-0.02 \mu\text{m}$  and standard

340 deviation of 0.56  $\mu\text{m}$ , whereas C99 model has a mean residual of -0.30  $\mu\text{m}$  and standard  
341 deviation of 0.58  $\mu\text{m}$ . The cause of this difference is made evident (Fig. 6A) by comparing the  
342  $l_{c,fit}$  vs.  $l_{a,fit}$  points against the lines that define their relationship (Donelick *et al.*, 1999, Equation  
343 1) in the two projection models. The B03 line passes through the center of the data, and the C99  
344 line intersects a cluster of points implying a steeper slope, or more quickly increasing anisotropy  
345 with increasing annealing.

346 To further examine the data, lines were fitted to the four  $l_{c,fit}$  vs.  $l_{a,fit}$  data points for each  
347 sample by each analyst. The resulting slope and intercept parameters are shown in Figure 6B.  
348 The B03 line parameters lie in the midst of the resulting point cluster, while the C99 parameters  
349 are close to the high-slope extreme, excluding outliers.

350 When deriving the two **c**-axis projection models, Ketcham *et al.* (2007a) postulated that  
351 their difference in slope may be due to the difference in etchant strength, with stronger etchant  
352 leading to higher anisotropy. This idea can be tested with the data in this study by observing  
353 how slope varies with etchant across the range used in this study (Fig. 6C). Overall, we find no  
354 clear signal; the range of slopes obtained for 5.0 M  $\text{HNO}_3$  encompasses the entire range observed  
355 for both 5.5 M  $\text{HNO}_3$  and also weaker etchants. The strongest etchant (7 M  $\text{HNO}_3$ ) appears to  
356 feature the highest slope, but those data are among the outliers. The slopes from the lab 13  
357 experiments testing various etchant strengths by the same analysts also show no clear pattern;  
358 analyst 1 got equivalent slopes for the strongest and weakest etchants, and the highest slope for  
359 the 5.0 M  $\text{HNO}_3$ , while analyst 2 got the lower slope for the 5.0 M and the highest for the  
360 weakest etchant. In part, these results reflect that the four-point line fits have considerable  
361 variability.

362 *Replicate analyses*

363 Replicate analyses by lab 32 showed no evidence of variation among aliquots. Although  
364 there was some divergence between answers beyond predicted statistical uncertainty, these were  
365 not systematic. For example, analyst 1 measured the longest mean track length on the third  
366 aliquot of sample DUR-2, whereas analyst 3 measured the shortest. On none of the four samples  
367 did they agree on the aliquot with the longest and shortest mean length. We thus conclude that  
368 variation among the apatite crystals in this study is at most a secondary effect.

#### 369 *TINT vs. TINCLE and $^{252}\text{Cf}$ irradiation*

370 We detected no indication in the reported data, particularly those for lab 14, that TINT  
371 and TINCLE measurements systematically diverge. Similarly, measurements obtained using  
372 exclusively tracks revealed by  $^{252}\text{Cf}$  irradiation by lab 41 analysts 2 and 5 showed no significant  
373 or systematic differences from other measurements by lab 41.

#### 374 *Normalization*

375 The large degree of variation observed among laboratories and analysts is likely to be due  
376 in part to persistent factors, such as laboratory instrumentation or procedures or systematic  
377 differences in analyst training or decision-making. We thus used the results for unannealed  
378 samples (DUR-2) for each analyst to normalize results for their annealed samples.

379 The results of two normalizations are shown in Figure 7. Normalizing based on the mean  
380 length of DUR-2 (Fig. 7A-C) considerably increases agreement among the data. The proportions  
381 of analyses within 2 standard errors of the overall mean increase to 48%, 37%, and 49% for  
382 DUR-4, DUR-1, and DUR-3, respectively. Convergence is similar and perhaps somewhat better  
383 when the track length data are c-axis projected with the B03-based model (Fig. 7D-F), with 42%,  
384 43%, and 52% of analyses within two standard errors of the mean. Though the comparison is

385 imperfect because not all analysts reported angle data, it is noteworthy that this improvement  
386 comes despite the smaller uncertainties of the c-axis projected means, which feature standard  
387 errors on average 29%, 33%, and 42% smaller for DUR-4, DUR-1, and DUR-3, respectively.

### 388 *Etch figures*

389 Figure 8 summarizes the measurements of etch figure long axis diameter ( $D_{par}$ ). As with  
390 the track length data, measurements for the unannealed sample show considerable variation.  
391 Interestingly, as with the track length data, this variation does not seem to correlate with etching  
392 procedure. If one considers the two principal protocols (5.0M and 5.5M HNO<sub>3</sub>, 20s, 21°C), it  
393 would be expected that the stronger etchant would result in larger etch figures (Sobel and  
394 Seward, 2010). However, the aggregate data do not show this (Fig. 8B).

395 Again, we normalized the  $D_{par}$  data for the annealed samples for each analyst using their  
396 respective measurements for sample DUR-2 (Fig. 8C). With the exception of some outliers, data  
397 for a given analyst are shown to be generally consistent to within  $\pm 10\%$ . The slightly lower  
398 normalized values for DUR-3 provide some indication that  $D_{par}$  may be influenced by annealing,  
399 although the effect is subtle. The lesser degrees of  $D_{par}$  shortening in the other annealed samples  
400 may be due to the less severe annealing conditions, or slight chemical variation; DUR-2 and  
401 DUR-3 are from same crystal, whereas the other experiments were from different crystals.

### 402 **Discussion**

403 Although concerns about the reproducibility of track-length data certainly arise from  
404 these experiments, the predominant picture has many positive aspects. Given that this  
405 experiment was used in part as a training aid by many laboratories (for example, as a benchmark  
406 for inexperienced analysts), full congruence of measurements is unrealistic. Length

407 measurements by all analysts arranged all samples into their correct ordering in term of  
408 annealing level, and the fact that the angular pattern observed by Carlson et al. (1999) and  
409 Barbarand et al. (2003a) is now repeated across many laboratories is an encouraging sign of  
410 consistency in the community. Also encouraging is that many of the differences among analysts  
411 are systematic enough that they can be substantially reduced by normalization with respect to a  
412 uniform standard of unannealed induced tracks.

413         It is important to note that there is no “correct” answer, as there will always be real  
414 differences due to etching, microscopy, and analyst decision-making. However, there are  
415 “incorrect” answers, which can be recognized as departures from the widely-observed patterns.  
416 These took a variety of forms: scatter at all angles (L05-A2, L21-A3, L25-A1, L32-A4);  
417 increased observation of short tracks across various angles (L38-A1 samples 1 and 4, 1-L21-A5,  
418 L28-A2, L14-A3 samples 2 and 4); scattered or short low-angle tracks (4-L21-A1); out-of-place  
419 short high-angle tracks (4-L12-A1).

420         A very interesting result is how non-influential etching is to the overall patterns in these  
421 data. There are no strong tendencies observed that can be traced to etching variations,  
422 particularly among the most commonly-employed protocols. This indicates that most variation  
423 present is due to the analyst rather than the etching procedure used. We stress, however, that we  
424 are not at all diminishing the importance of strict attention to detail when etching; this study was  
425 not designed to test the consequences of poor etching procedure.

## 426 **Recommendations**

427         The two most crucial lessons that arise from the results presented here are the importance  
428 of normalization and training. Also of interest is the optimal method for c-axis projection among  
429 laboratories that use it. We present below recommendations for each of these.

430 *Normalization*

431 All length measurements should be normalized before interpretation using thermal  
432 history inverse modeling, to ensure that they are compatible with the measurements underlying  
433 the annealing models. As a minimum step, initial length track should be normalized to Durango  
434 apatite, either using sample DUR-2 from this study or an independently-created induced-track  
435 sample. However, insofar as initial induced track length is known to vary with apatite chemistry  
436 and solubility (Carlson *et al.*, 1999), a more thorough procedure that takes this into account is  
437 preferable. The measurable parameter that is best-correlated with initial track length is  $D_{par}$   
438 (Carlson *et al.*, 1999). Four methods might thus be considered:

439 1) *Use the DUR-2 measurement as the “initial track length” for modeling software.*

440 This is at best a first-order correction, as it neglects that Durango apatite actually has  
441 a slightly longer initial track length than typical F-apatites (Carlson *et al.*, 1999),  
442 which are the most commonly-encountered variety in practice.

443 2) *Adjust length measurements using DUR-2, and use published models for*  
444 *extrapolation.* An adjustment factor for mean length,  $a_{lm}$ , based on DUR apatite can  
445 be calculated for an analyst as:

446 
$$a_{lm} = \frac{l_{m,DUR,published}}{l_{m,DUR,analyst}} \quad (1)$$

447 where the numerator is the unannealed induced Durango mean track length  
448 measurement underlying the published annealing model calibration (i.e. from Carlson  
449 *et al.*, 1999 or Barbarand *et al.*, 2003) and the denominator is a particular analyst’s  
450 corresponding measurement. Length measurements can be multiplied by this factor  
451 before being entered into modeling software, or the software may allow entry of  $a_{lm}$ .  
452 It carries the advantage of still using etch figures or composition (assuming they are



453 measured) to better approximate initial length, and leverages the many measurements  
454 that underlie the published calibrations. If etch figures are used, they would require a  
455 similar adjustment factor:

$$456 \quad a_{Dpar} = \frac{D_{par,DUR,published}}{D_{par,DUR,analyst}} \quad (2)$$

457 The primary shortcoming of this solution is that it is based on a single measurement,  
458 which provides only limited information on whether there is a change in how length  
459 varies among apatites.

460 3) *Use the same method as 2, with more apatite varieties.* Sobel and Seward (2010)  
461 advocate a cross-calibration of  $D_{par}$  data using two apatite standards, Durango and  
462 Fish Canyon, in which the user-measured values are plotted against the published  
463 ones, and a line is fitted through them which also passes through the origin. This  
464 approach is mathematically equivalent to option 2 above, simply averaging together  
465 two or even more apatites, and can be applied equally to length data. Thus:

$$466 \quad a_{lm} = \frac{\sum l_{m,published}}{\sum l_{m,analyst}} \quad (3)$$

$$467 \quad a_{Dpar} = \frac{\sum D_{par,published}}{\sum D_{par,analyst}} \quad (4)$$

468 This approach has the advantage of being less sensitive to a single analysis, and  
469 incorporating information from different apatites.

470 4) *Construct complete new calibrations between initial track length and solubility or*  
471 *composition using multiple apatites.* This method would be most rigorous, but also  
472 the most demanding, both in terms of effort and the demands placed upon the  
473 experimental material. In particular, it would be necessary to have samples spanning

474 the range of solubility/composition, as well as the variability in the initial track length  
475 documented in F-apatites (Carlson et al., 1999).

476 Figure 9 illustrates normalization methods 1 through 3, using data measured by an analyst  
477 for Durango and Fish Canyon standards (DUR  $l_m=16.05\pm0.08$   $\mu\text{m}$ ,  $D_{par} = 1.98\pm0.03$   $\mu\text{m}$ ; FCT  
478  $l_m=16.05\pm0.08$   $\mu\text{m}$ ,  $D_{par} = 2.44\pm0.04$   $\mu\text{m}$ ). In this example it is assumed that the analyst has  
479 decided that the C99 c-axis projection model is more appropriate, and thus the measurements  
480 should be normalized based on Carlson et al. (1999) data. The track length measurements are  
481 systematically lower than the corresponding ones from Carlson et al. (1999), but the  $D_{par}$   
482 measurements are slightly higher (DUR  $l_m=16.21\pm0.08$   $\mu\text{m}$ ,  $D_{par} = 1.83\pm0.03$   $\mu\text{m}$ ; FCT  
483  $l_m=16.38\pm0.08$   $\mu\text{m}$ ,  $D_{par} = 2.43\pm0.04$   $\mu\text{m}$ ). Figure 9 shows the Carlson et al. (1999)  $l_m$  and  $D_{par}$   
484 data, with apatites DUR and FCT highlighted, along with the published linear fit. The analyst's  
485 corresponding measurements are plotted, as well as the linear relationships based on each  
486 normalization method. Method 1 provides an invariant line, which is sub-optimal but arguably  
487 defensible if only near-end-member F-apatites are being considered. Method 2 captures the  
488 variation in initial length documented by Carlson et al. (1999), but with a slightly different slope  
489 caused by the 7.6% increase in  $D_{par}$  values ( $a_{l_m} = 1.010$ ,  $a_{D_{par}} = 0.924$ ). When results for FCT  
490 apatites are averaged in (method 3), the slope becomes more similar to the published one due to  
491 the  $D_{par}$  increase being reduced to 3.4%, which is further offset by the  $l_m$  decrease of 1.1% ( $a_{l_m} =$   
492  $1.011$ ,  $a_{D_{par}} = 0.966$ ).

493 Of the options discussed, our recommendation is that analysts use method 3, or otherwise  
494 method 2 if further annealed standards with induced tracks cannot be obtained. These options  
495 leverage the large amount of existing calibration data that underlie the published relationships  
496 between initial length and solubility or composition, which makes them more likely to give

497 reasonable answers when applied to unusual apatite varieties (i.e. large etch figures). Although  
498 the difference between methods 2 and 3 is minor in the example shown in Figure 9, we feel that,  
499 analogously to age zeta calibration (Hurford, 1990), best practice requires evaluation of multiple  
500 standards. The effort required for option 4 is probably only justified if the etching protocol is  
501 severely changed, such as by using a weak etchant and/or substantially longer etching times.

502 A full analysis of the ramifications of neglecting normalization for inverse modeling are  
503 beyond the scope of this study, as they are very context-dependent based on the types of samples  
504 and geological histories being investigated. Ketcham et al. (2009) showed that omitting  
505 normalization affected the shape of the fitted cooling path and the final cooling temperature in  
506 cooling-only histories, and the maximum reheating temperature in non-monotonic histories.

#### 507 *C-axis projection*

508 C-axis projection seems to increase inter-laboratory compatibility, and accounts for some  
509 differing operator tendencies, particularly at strong levels of annealing. It also removes a  
510 substantial component of noise: all tracks at a given level of annealing (i.e. Fig. 5) reflect the  
511 same thermal input despite their difference in length, and c-axis projection responds by utilizing,  
512 rather than discarding, the information in track angle.

513 It is worth reconsidering which is the appropriate c-axis projection model for a given  
514 analyst or lab group to use. The B03 model seems to represent the majority of the community,  
515 but some labs are better represented by C99 model. The spread in Figure 6B indicates that four-  
516 point fits are not enough data to make a definitive judgment in any single case, however. The  
517 C99 model tends to result in larger  $l_{c,mod}$  values, because increased anisotropy means that the  
518 lengths of high- $\phi$  tracks are increased more when being projected to c-axis-parallel. Thus, in  
519 Figure 7D-F, the C99 projection model will cause the data for a given lab or analyst to move

520 rightwards, slightly at low degrees of annealing and more at higher degrees. In some cases this  
521 step can increase compatibility among analysts and lab groups; for example, using C99 for lab 13  
522 appears on the whole to increase intercompatibility with other large lab groups (14, 32).

523         It is in fact possible to create a “tunable” parameter that adjusts the projection model  
524 slope+intercept to maximize compatibility among laboratories. However, we are cautious about  
525 recommending such a step, as we have not ascertained the reasons underlying this apparent  
526 change in the evolution of anisotropy among analysts. The divergent answers in C99 and B03  
527 are both real, as they each reflect dozens of careful experiments, but we don’t yet know what  
528 makes them real. Also, again, some divergences observed in this study are likely to be simply a  
529 case of inadequate training or attention to detail, and it would be unwise to create a fudge factor  
530 to compensate for this rather than addressing the root of the problem.

531         We thus recommend that each lab evaluate for itself which is the preferable model to  
532 employ, using the four samples from this study, or equivalents, to decide which more closely  
533 reflects the measurements they produce. The model chosen should be reported when thermal  
534 history inversion is used.

### 535 *Training*

536         In addition to the protocols developed by experienced workers for their respective  
537 laboratories, the samples distributed in this study can be a useful training resource. In particular,  
538 we recommend a training regimen of measuring these four samples (or equivalents) and  
539 critiquing the results, if necessary repeatedly, until they are judged suitably compatible with the  
540 community, before measurement and utilization of track lengths for research is undertaken.  
541 Results should be evaluated not only for the usual mean and standard deviation but also for

542 consistent distribution with angle, and careful attention to how borderline cases are evaluated and  
543 pitfalls avoided.

544           These mounts are good for training and inter-laboratory comparison, but not perfect.  
545 They are simple length-angle distributions, and after an analyst begins to measure a pattern may  
546 be recognized, which can in turn bias further measurements of that sample. Measuring “blind”  
547 (i.e. not seeing tables or summaries of measurements as they are made) is thus crucial. The use of  
548 megacryst slabs does not test for grain selection, and the plentitude of tracks may shift an  
549 analyst’s bias in evaluating track suitability (i.e. borderline cases may be more likely to be  
550 passed over). Measurement of the samples in this study also provides no information about  
551 relative probability of sampling from different length populations, and whether this tendency  
552 varies among analysts. In view of these considerations, samples containing multiple, known  
553 annealed populations would be a valuable additional inter-calibration and training tool, and are  
554 being created by the lead author.

## 555 **Implications**

556           The results of this study indicate that there has been some degree of scatter in apatite  
557 fission-track length data used for research that can be traced to the human element in their  
558 measurement, which in turn is likely to affect some aspects of thermal history inverse modeling.  
559 It is also clear that a large component of the problem can be addressed fairly simply and easily  
560 through normalizing for length and angle and, where necessary, enhancing training regimens.

561           The time is approaching when automated systems may take over the measurement of  
562 length data, which may enhance or even come close to assuring inter-laboratory consistency.  
563 Even over the interval between when most of the measurements in this study were made and the  
564 present, there has been considerable progress in using image-analysis-based methods to improve

565 the picking of track end-points (Donelick et al., 2013; Gleadow and Seiler, 2013). However, full  
566 automation of track identification and evaluation is still some time away, with universal  
567 acceptance and adoption of these solutions even further in the future. In the intervening time, the  
568 measures advocated here should serve to improve the overall quality of length data produced by  
569 the fission-track community, and in turn the interpretations generated from those data.

## 570 **Acknowledgements**

571 Our foremost and deepest thanks go to the many individuals who participated in this  
572 study. Constructive reviews by R. Flowers and K. Gallagher improved the manuscript. We also  
573 thank the Jackson School of Geosciences for its support of the University of Texas fission-track  
574 laboratory, and M. Tamer for helping draft figure 1 and contributing suggestions and additional  
575 measurements.

## 576 **References**

- 577 Barbarand, J., Carter, A., Wood, I., and Hurford, A.J. (2003a) Compositional and structural control of  
578 fission-track annealing in apatite. *Chemical Geology*, 198, 107-137.
- 579 Barbarand, J., Hurford, A.J., and Carter, A. (2003b) Variation in apatite fission-track length measurement:  
580 implications for thermal history modelling. *Chemical Geology (Isotope Geoscience Section)*,  
581 198, 77-106.
- 582 Bhandari, N., Bhat, S.G., Lal, D., Rajagopalan, G., Tamhane, A.S., and Venkatavaradan, V.S. (1971) Fission  
583 fragment tracks in apatite: recordable track lengths. *Earth and Planetary Science Letters*, 13,  
584 191-199.
- 585 Burtner, R.L., Nigrini, A., and Donelick, R.A. (1994) Thermochronology of lower Cretaceous source rocks  
586 in the Idaho-Wyoming thrust belt. *American Association of Petroleum Geologists Bulletin*,  
587 78(10), 1613-1636.
- 588 Carlson, W.D., Donelick, R.A., and Ketcham, R.A. (1999) Variability of apatite fission-track annealing  
589 kinetics I: Experimental results. *American Mineralogist*, 84, 1213-1223.
- 590 Crowley, K.D., Cameron, M., and Schaefer, R.L. (1991) Experimental studies of annealing etched fission  
591 tracks in fluorapatite. *Geochimica et Cosmochimica Acta*, 55, 1449-1465.
- 592 Donelick, R.A. (1991) Crystallographic orientation dependence of mean etchable fission track length in  
593 apatite: An empirical model and experimental observations. *American Mineralogist*, 76, 83-91.
- 594 Donelick, R.A., Donelick, M.B., O'Sullivan, P.B., McMillan, J., Hourigan, J.K., and Juel, E. (2013) Three-  
595 dimensional spatial characteristics and contents of zircon crystals from high resolution optical  
596 imagery for the fission track, (U-Th-Sm)/He, and U-Th-Pb systems. *American Geophysical Union*  
597 *Fall Meeting*, p. T42C-06, San Francisco.

598 Donelick, R.A., Ketcham, R.A., and Carlson, W.D. (1999) Variability of apatite fission-track annealing  
599 kinetics II: Crystallographic orientation effects. *American Mineralogist*, 84, 1224-1234.

600 Donelick, R.A., and Miller, D.A. (1990) Enhanced TINT fission track densities in low spontaneous track  
601 density apatites using <sup>252</sup>Cf-derived fission fragment tracks: A model and experimental  
602 observations. *International journal of radiation applications and instrumentation. Part D,*  
603 *Nuclear tracks and radiation measurements*, 18(3), 301-307.

604 Donelick, R.A., O'Sullivan, P.B., and Ketcham, R.A. (2005) Apatite fission-track analysis. In P.W. Reiners,  
605 and T.A. Ehlers, Eds. *Reviews in Mineralogy and Geochemistry*, 58, p. 49-94.

606 Galbraith, R.F. (2005) *Statistics for Fission Track Analysis*. 219 p. Chapman and Hall, Boca Raton.

607 Galbraith, R.F., Laslett, G.M., Green, P.F., and Duddy, I.R. (1990) Apatite fission track analysis: geological  
608 thermal history analysis based on a three-dimensional random process of linear radiation  
609 damage. *Philosophical Transactions of the Royal Society of London A*, 332, 419-438.

610 Gallagher, K. (1995) Evolving temperature histories from apatite fission-track data. *Earth and Planetary*  
611 *Science Letters*, 136, 421-435.

612 Gallagher, K. (2012) Transdimensional inverse thermal history modeling for quantitative  
613 thermochronology. *Journal of Geophysical Research*, 117, B02408.

614 Gleadow, A.J.W., Duddy, I.R., Green, P.F., and Lovering, J.F. (1986) Confined fission track lengths in  
615 apatite: a diagnostic tool for thermal history analysis. *Contributions to Mineralogy and*  
616 *Petrology*, 94, 405-415.

617 Gleadow, A.J.W., and Seiler, C. (2013) Fission track length distributions in multi-system  
618 thermochronology. *American Geophysical Union Fall Meeting*, p. T42C-08, San Francisco.

619 Green, P.F., Duddy, I.R., Gleadow, A.J.W., Tingate, P.R., and Laslett, G.M. (1986) Thermal annealing of  
620 fission tracks in apatite 1. A qualitative description. *Chemical Geology (Isotope Geoscience*  
621 *Section)*, 59, 237-253.

622 Green, P.F., Duddy, I.R., Laslett, G.M., Hegarty, K.A., Gleadow, A.J.W., and Lovering, J.F. (1989) Thermal  
623 annealing of fission tracks in apatite 4. Quantitative modeling techniques and extension to  
624 geological time scales. *Chemical Geology (Isotope Geoscience Section)*, 79, 155-182.

625 Green, P.F., and Durrani, S.A. (1977) Annealing studies of tracks in crystals. *Nuclear Track Detection*,  
626 1(1), 33-39.

627 Hurford, A.J. (1990) Standardization of fission track dating calibration: Recommendation by the Fission  
628 Track Working Group of the I.U.G.S. Subcommittee on Geochronology. *Chemical Geology*  
629 *(Isotope Geoscience Section)*, 80, 171-178.

630 Jonckheere, R., Enkelmann, E., Min, M., Trautmann, C., and Ratschbacher, L. (2007) Confined fission  
631 tracks in ion-irradiated and step-etched prismatic sections of Durango apatite. *Chemical*  
632 *Geology*, 242, 202-217.

633 Jonckheere, R., and Ratschbacher, L. (2010) On measurements of non-horizontal confined fission tracks.  
634 In R.W. Brown, Ed. *12th International Conference on thermochronometry*, p. 60, Glasgow.

635 Jonckheere, R., and Wagner, G.A. (2000) On the occurrence of anomalous fission tracks in apatite and  
636 titanite. *American Mineralogist*, 85, 1744-1753.

637 Ketcham, R.A. (2003) Observations on the relationship between crystallographic orientation and biasing  
638 in apatite fission-track measurements. *American Mineralogist*, 88, 817-829.

639 Ketcham, R.A. (2005) Forward and inverse modeling of low-temperature thermochronometry data. In  
640 P.W. Reiners, and T.A. Ehlers, Eds. *Reviews in Mineralogy and Geochemistry*, 58, p. 275-314.

641 Ketcham, R.A., Carter, A.C., Donelick, R.A., Barbarand, J., and Hurford, A.J. (2007a) Improved  
642 measurement of fission-track annealing in apatite using c-axis projection. *American*  
643 *Mineralogist*, 92, 789-798.

644 Ketcham, R.A., Carter, A.C., Donelick, R.A., Barbarand, J., and Hurford, A.J. (2007b) Improved modeling  
645 of fission-track annealing in apatite. *American Mineralogist*, 92, 799-810.

- 646 Ketcham, R.A., Donelick, R.A., Balestrieri, M.L., and Zattin, M. (2009) Reproducibility of apatite fission-  
647 track length data and thermal history reconstruction. *Earth and Planetary Science Letters*, 284,  
648 504-515.
- 649 Ketcham, R.A., Donelick, R.A., and Carlson, W.D. (1999) Variability of apatite fission-track annealing  
650 kinetics III: Extrapolation to geological time scales. *American Mineralogist*, 84, 1235-1255.
- 651 Lal, D., Rajan, R.S., and Tamhane, A.S. (1969) Chemical composition of nuclei of  $Z > 22$  in cosmic rays  
652 using meteoric minerals as detectors. *Nature*, 221, 33-37.
- 653 Laslett, G.M., and Galbraith, R.F. (1996) Statistical modelling of thermal annealing of fission tracks in  
654 apatite. *Geochimica et Cosmochimica Acta*, 60, 5117-5131.
- 655 Laslett, G.M., Green, P.F., Duddy, I.R., and Gleadow, A.J.W. (1987) Thermal annealing of fission tracks in  
656 apatite 2. A quantitative analysis. *Chemical Geology (Isotope Geoscience Section)*, 65, 1-13.
- 657 Laslett, G.M., Kendall, W.S., Gleadow, A.J.W., and Duddy, I.R. (1982) Bias in measurement of fission-  
658 track length distributions. *Nuclear Tracks and Radiation Measurements*, 6(2/3), 79-85.
- 659 Li, N., Wang, L., Sun, K., Lang, M., Trautmann, C., and Ewing, R.C. (2010) Porous fission fragment tracks in  
660 fluorapatite. *Physical Review B*, 82, 144109.
- 661 Miller, D.S., Crowley, K.D., Dokka, R.K., Galbraith, R.F., Kowallis, B.J., and Naeser, C.W. (1993) Results of  
662 interlaboratory comparison of fission track ages for 1992 Fission Track Workshop. *Nuclear*  
663 *Tracks and Radiation Measurements*, 21(4), 565-573.
- 664 O'Sullivan, P.B., Donelick, R.A., and Ketcham, R.A. (2004) Etching conditions and fitting ellipses: what  
665 constitutes a proper apatite fission-track annealing calibration measurement? In P. Andriessen,  
666 Ed. 10th International Conference on Fission Track Dating and Thermochronology, p. Abstract  
667 code DV-10-O, Amsterdam.
- 668 Sobel, E.R., and Seward, D. (2010) Influence of etching conditions on apatite fission-track etch pit  
669 diameter. *Chemical Geology*, 271, 59-69.
- 670 Yamada, R., Tagami, T., and Nishimura, S. (1995) Confined fission-track length measurement of zircon:  
671 assessment of factors affecting the paleotemperature estimate. *Chemical Geology (Isotope*  
672 *Geoscience Section)*, 119, 293-306.

673

674

675



## Figure Captions

Figure 1. Photomicrographs of confined fission tracks. Black scale bars are 10  $\mu\text{m}$ , and apatite  $c$  axis orientations marked with white arrows. (A) Transmitted light image of measurable track at intermediate angle to  $c$  axis. (B) Reflected light image of track in A. (C) Transmitted light image of tracks at  $\sim 25^\circ$  to  $c$  axis; track 1 is measurable, but track 2 intersects surface. (D) Reflected light image of field of view in C, also showing elongated etch figures indicating  $c$  axis direction. (E) Reflected light image with track 1 near parallel to  $c$  axis and track 2 near perpendicular. (F) Reflected light image with  $c$ -axis-perpendicular track with pinched ends. Images in (A-F) all obtained with transmitted light. (G) Reflected light image of track that appears shortened due to fluid. (H) Same track after fluid has been removed with acetone wash.

Figure 2. Mean track lengths and errors (1 SE) for unannealed sample DUR-2 versus: (A) lab code; (B) years since being trained in fission-track analysis as of the time study measurements were made; (C) approximate number of fission-track mounts measured per year over the previous 3 years; (D) etching method.

Figure 3. Mean track lengths and errors (1 SE) for annealed samples.

Figure 4. Polar plots of fission-track length measurements of four study samples from an experienced analyst (A-D) and a novice (E-H). Codes refer to sample number (i.e. 1-4 indicates DUR-1 through DUR-4), lab number, and analyst number.

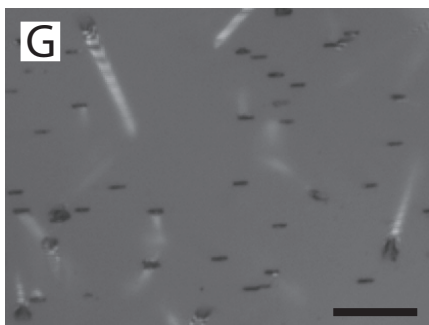
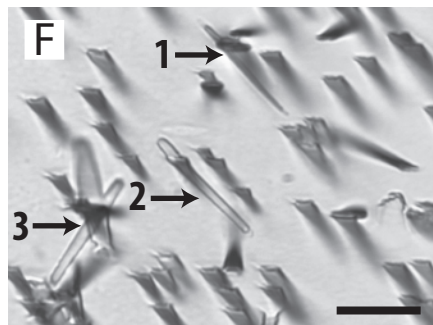
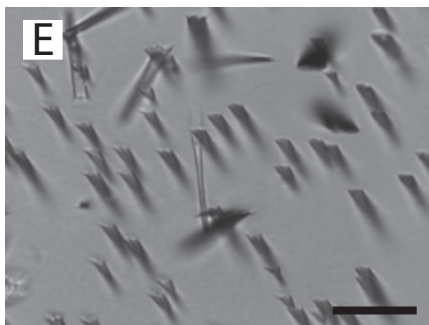
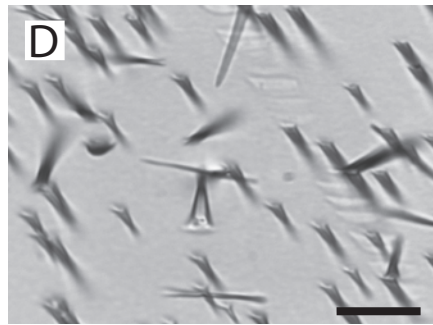
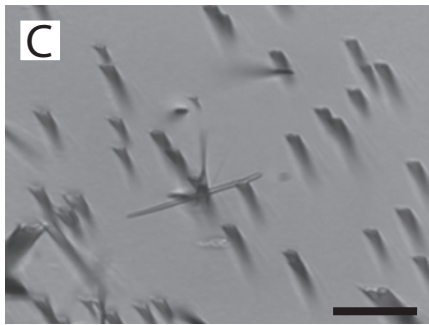
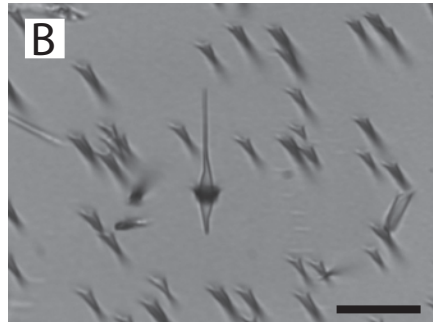
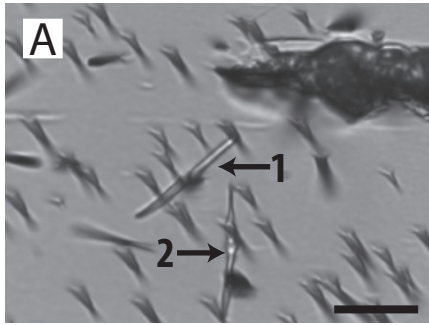
Figure 5. Polar plots of data for six experienced analysts for aliquots of sample DUR-3, showing different tendencies for measuring shortened tracks at high angles to the  $c$  axis.

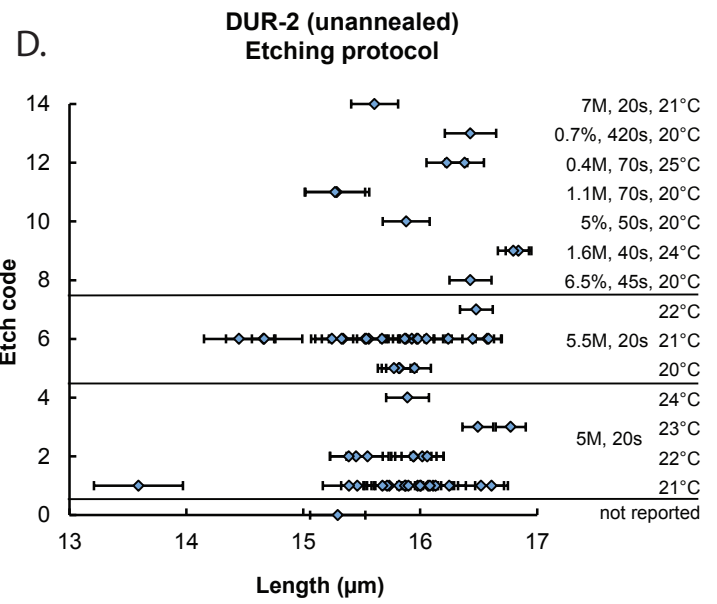
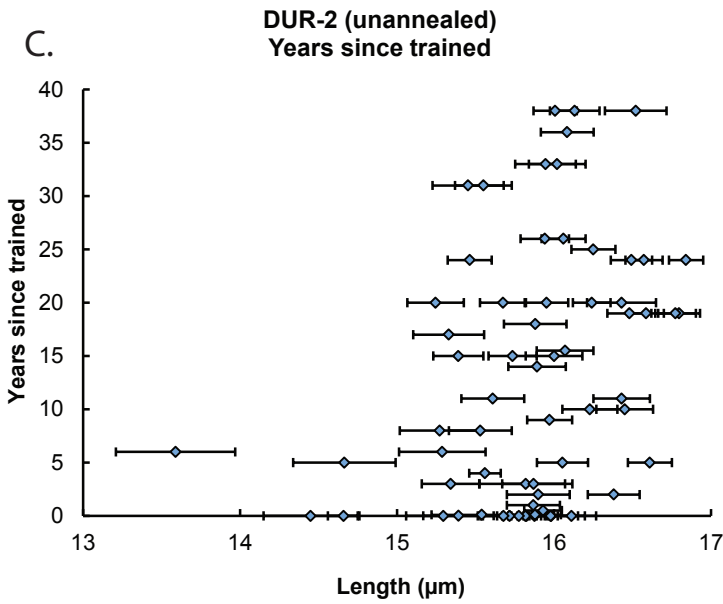
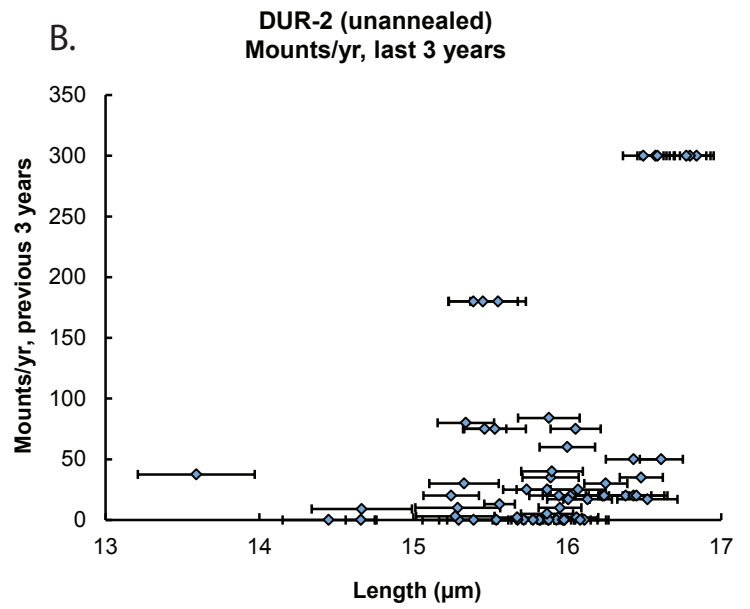
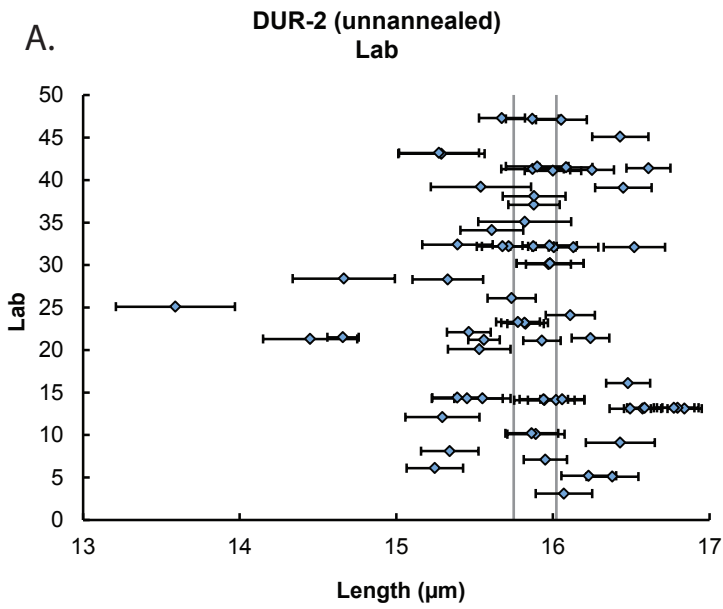
Figure 6. Summary plots showing evolution of track length anisotropy. (A) Individual  $l_{c,fit}$  vs.  $l_{a,fit}$  data, with lines representing this relationship from Ketcham et al. (2007a) based on data from Carlson et al. (1999) and Barbarand et al. (2003) (C99 and B03, respectively). (B) Slope and intercept of lines fit to four samples for each study participant, and corresponding points from C99 and B03. (C) Range of fitted  $l_{c,fit}$  vs.  $l_{a,fit}$  slopes for each etching method reported.

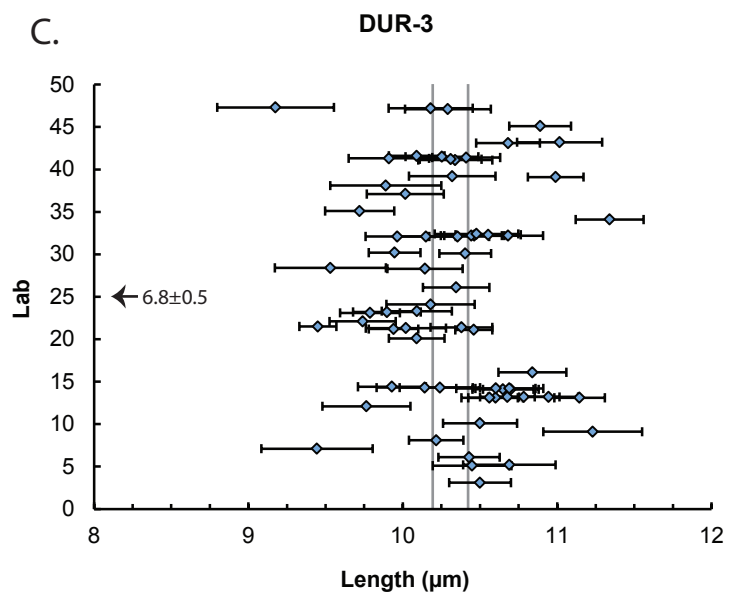
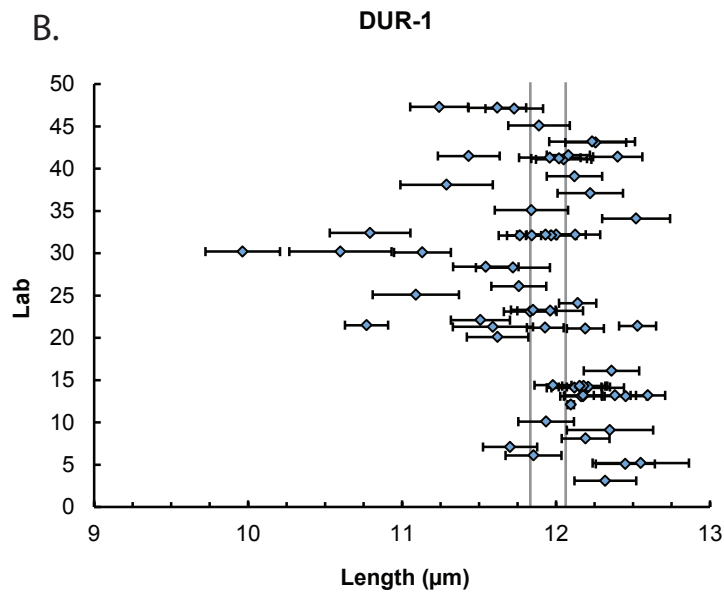
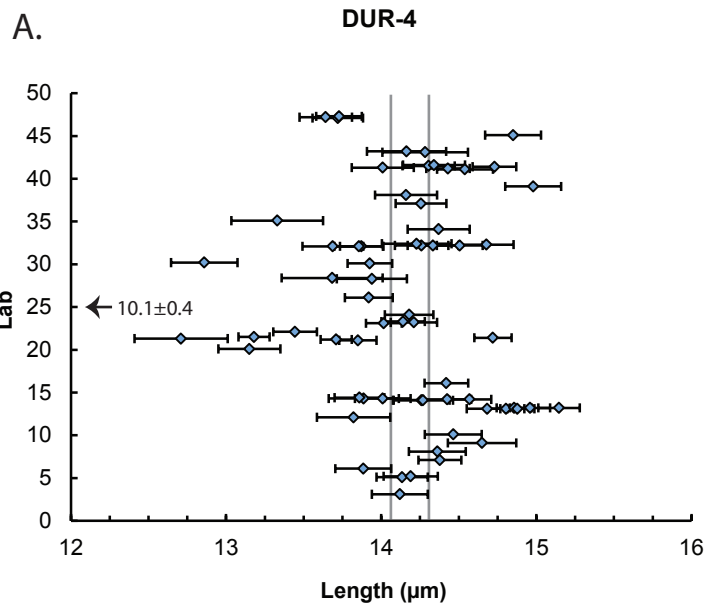
Figure 7. Normalized lengths and errors (1 SE) for annealed samples, for mean (A-C) and  $c$ -axis-projected (D-F) data.

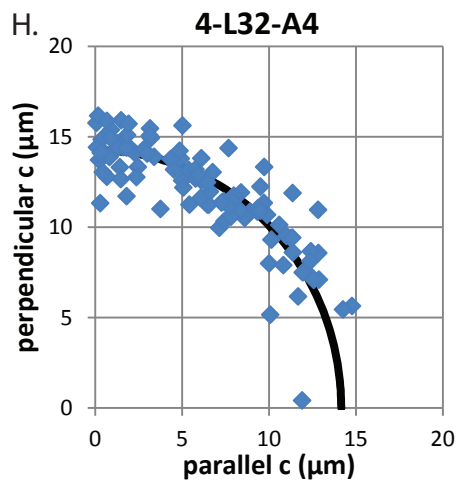
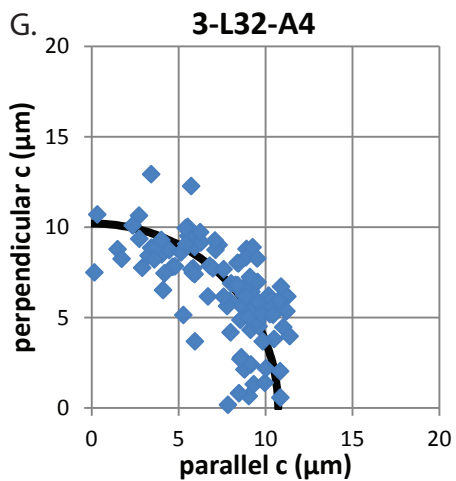
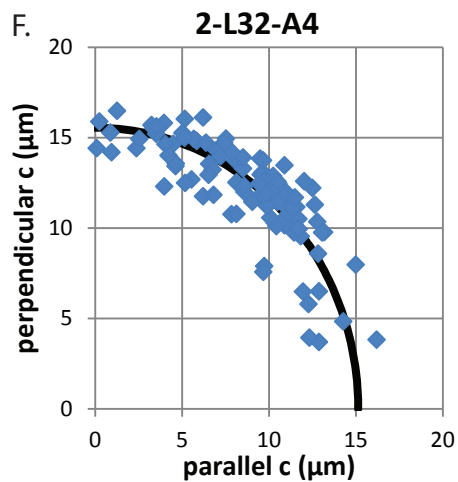
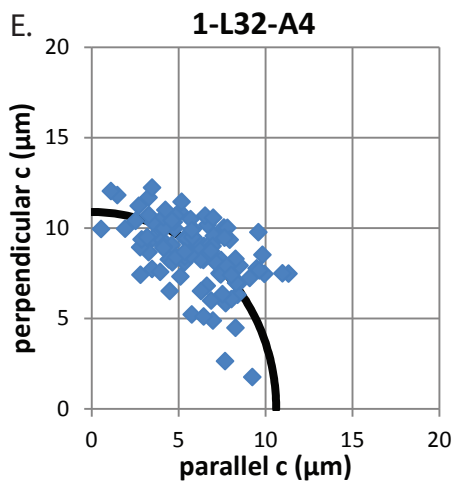
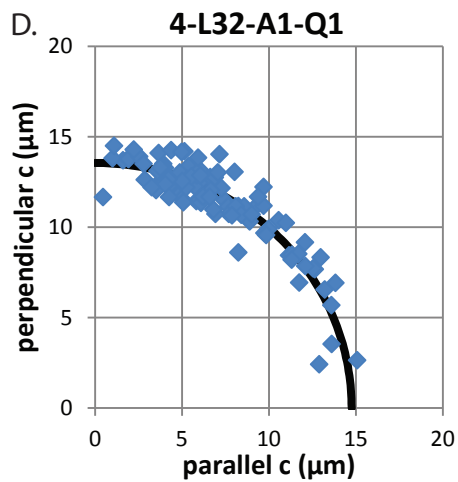
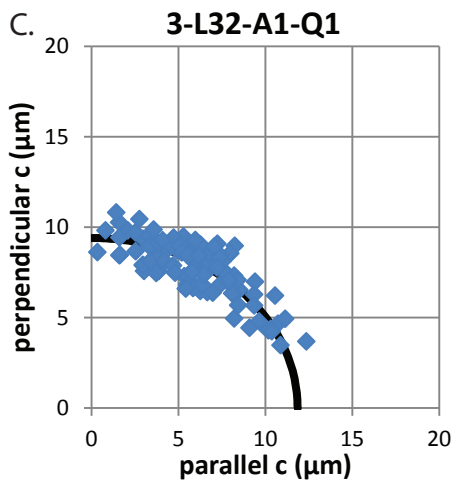
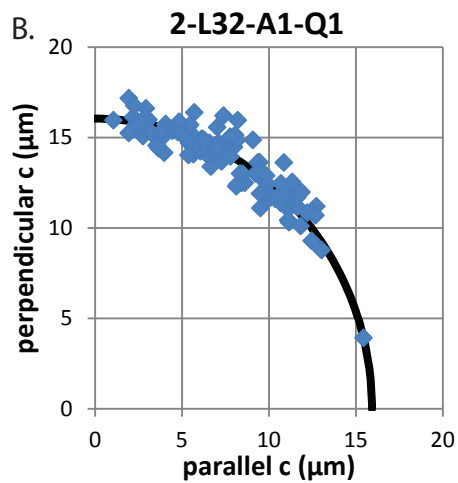
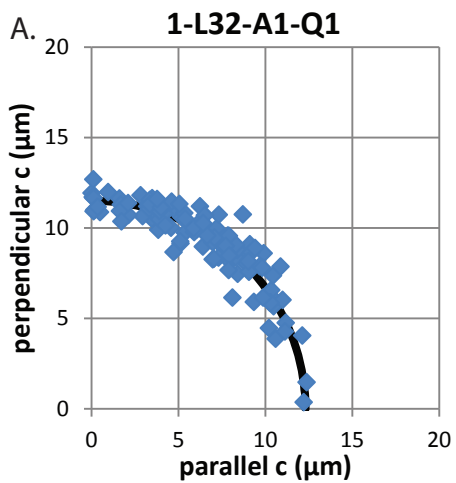
Figure 8. Summary plots of etch figure length ( $D_{par}$ ) data. (A) Mean  $D_{par}$  and error (1 SE) versus lab number. (B) Mean  $D_{par}$  and error versus etching method. (C) Normalized  $D_{par}$  values for each sample.

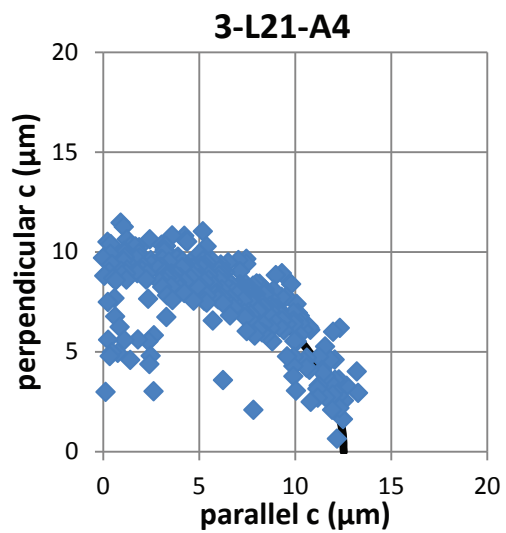
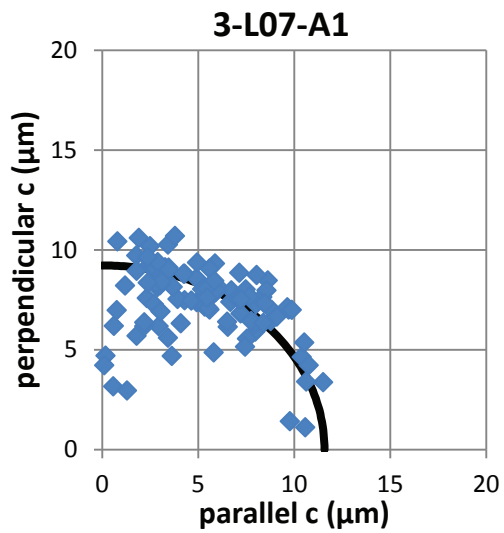
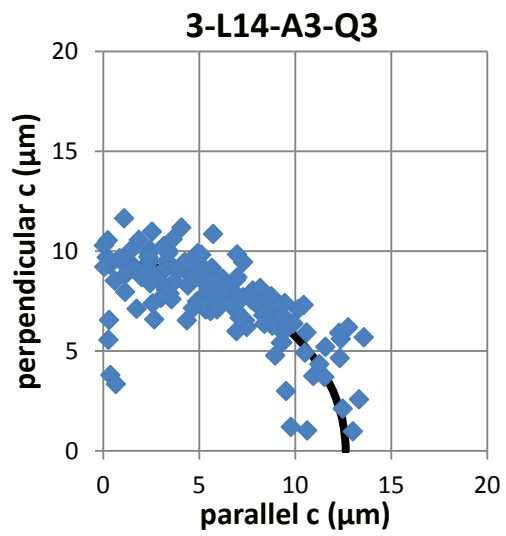
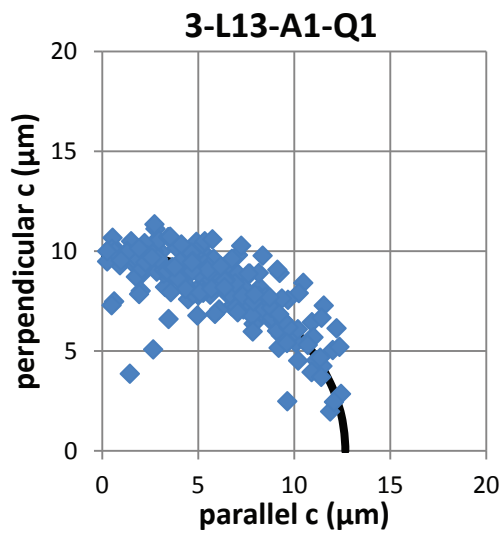
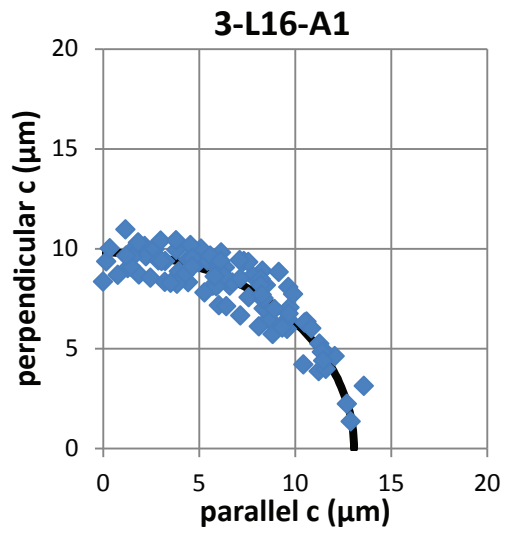
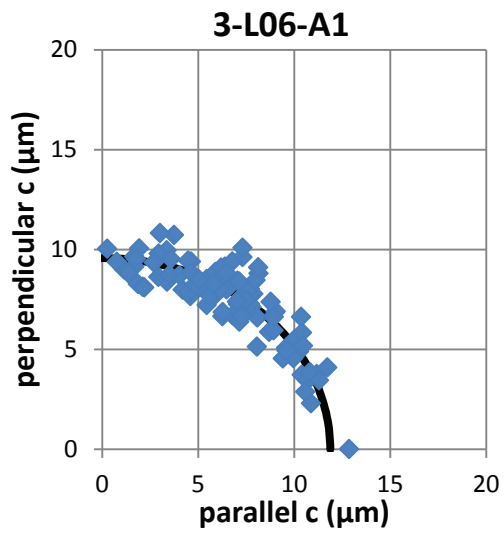
Figure 9. Example showing outcome of three normalization methods for track length and etch figure data discussed in text.

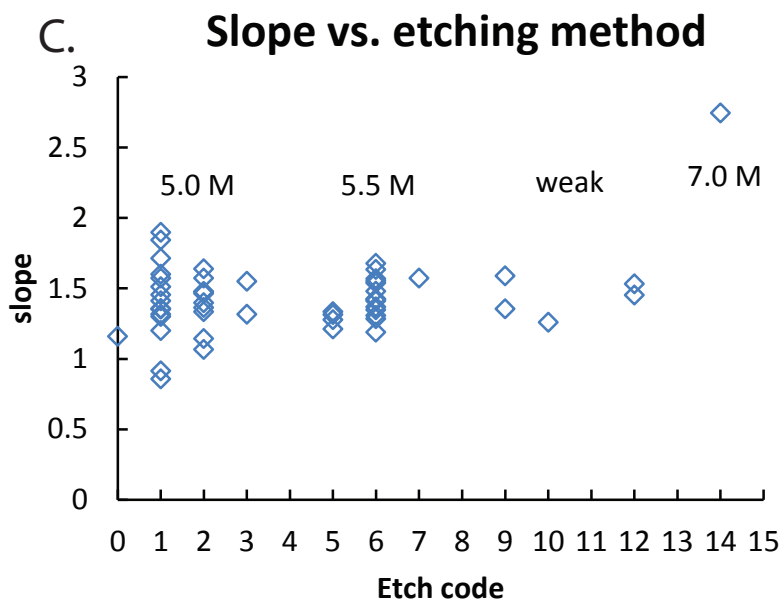
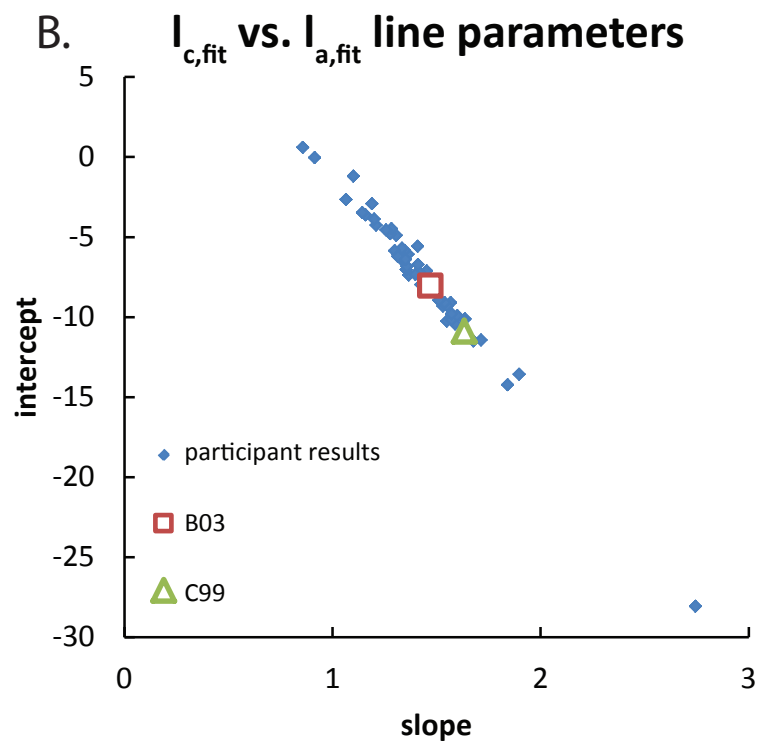
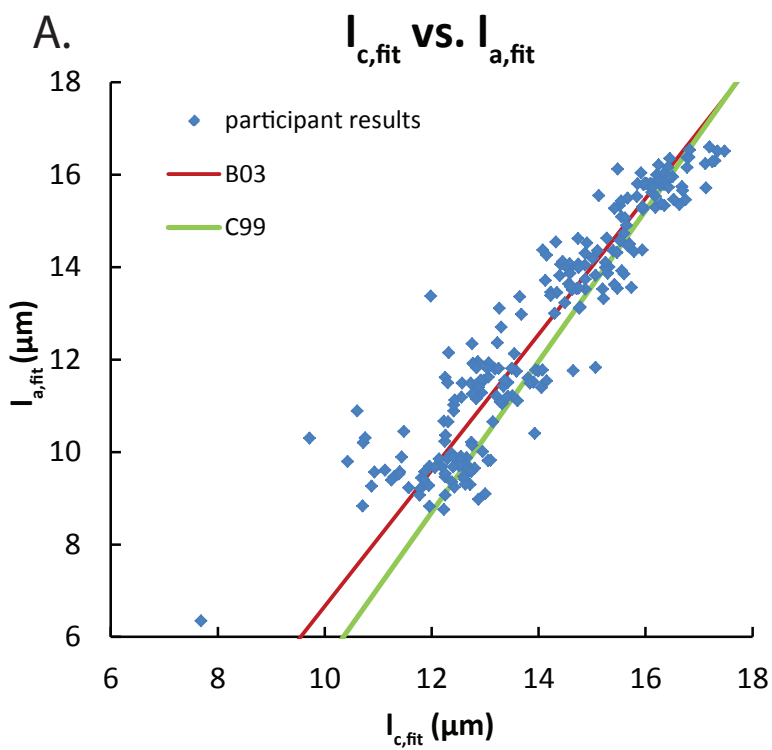


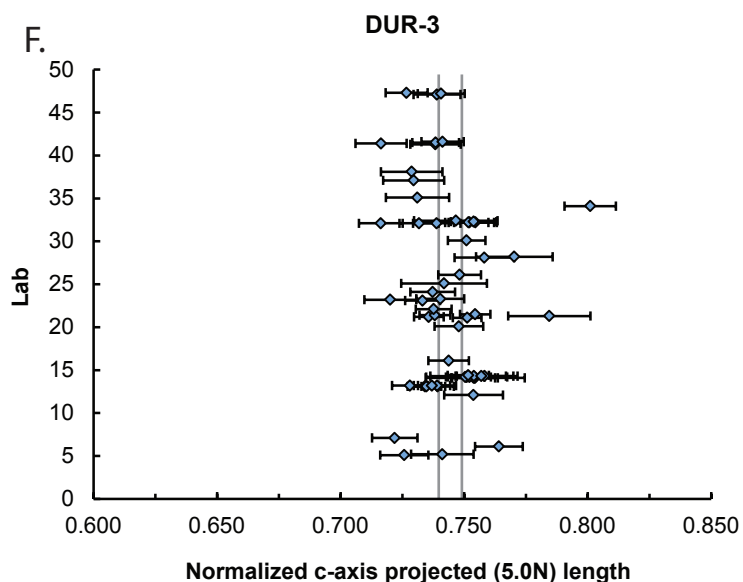
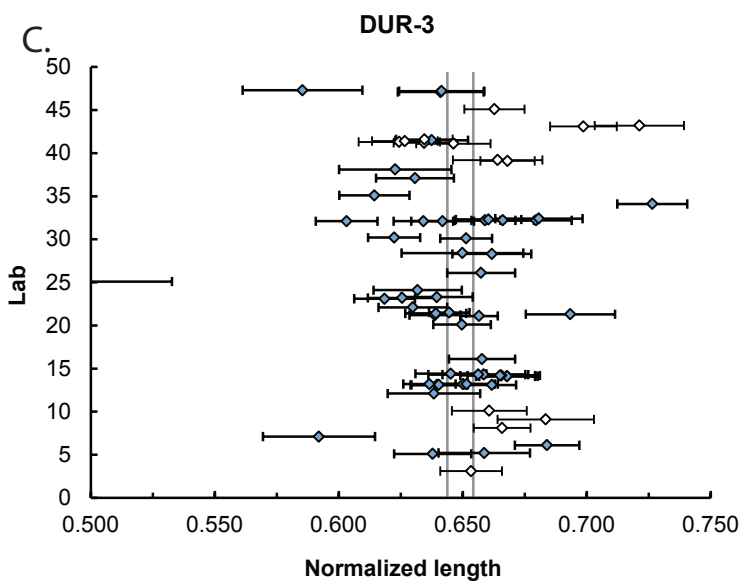
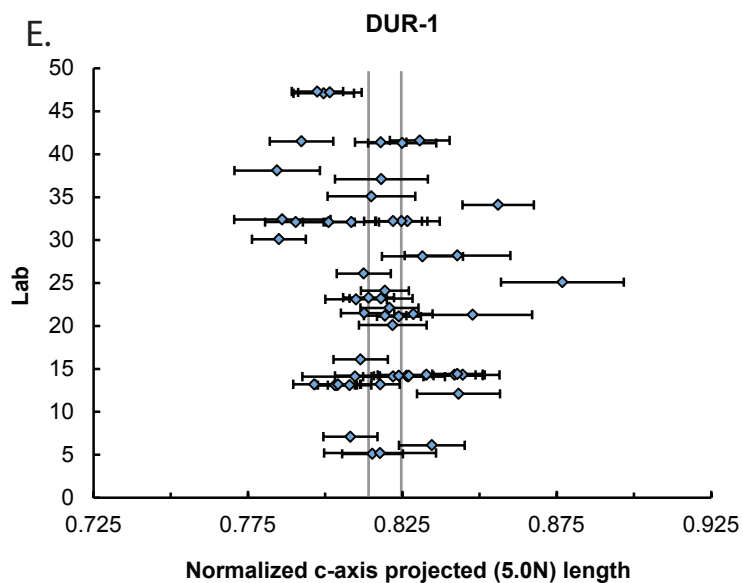
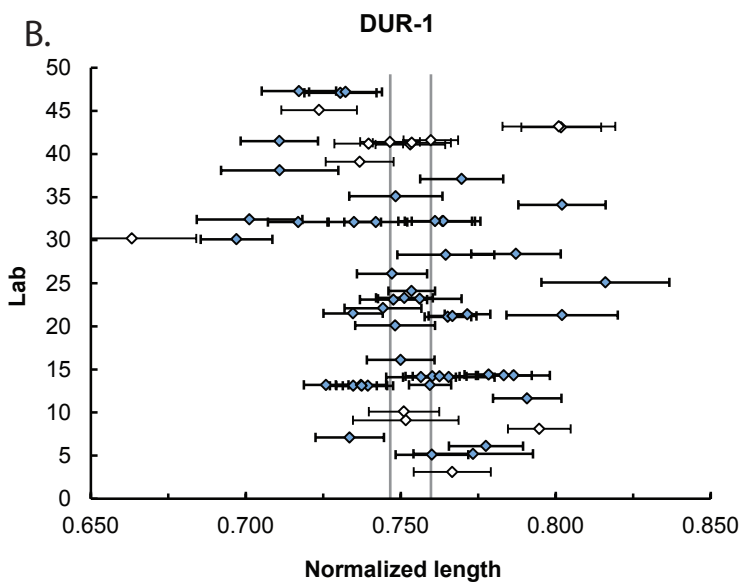
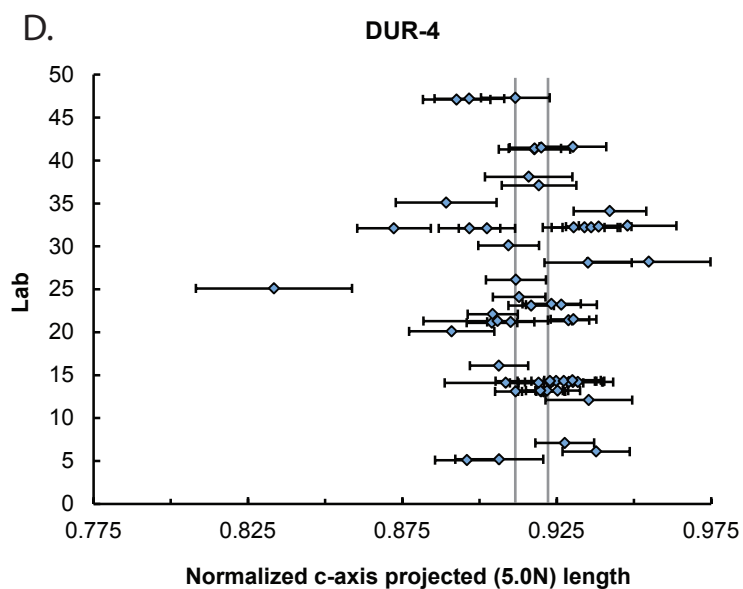
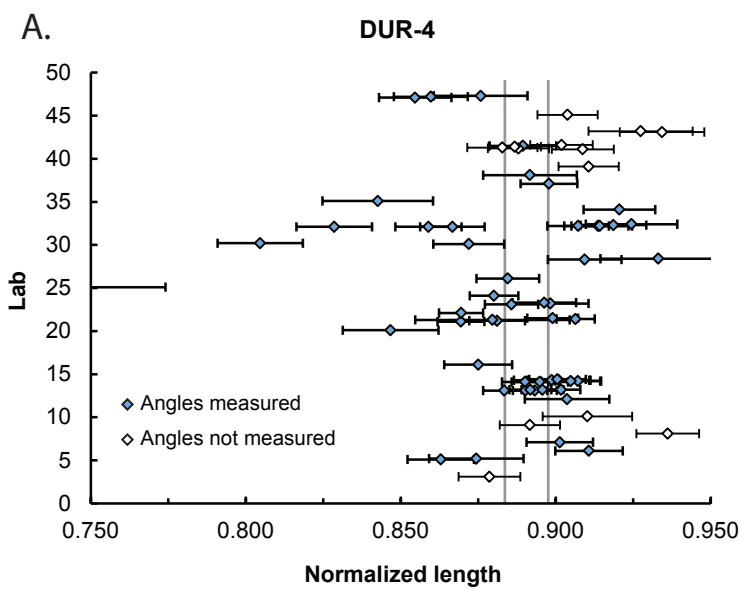




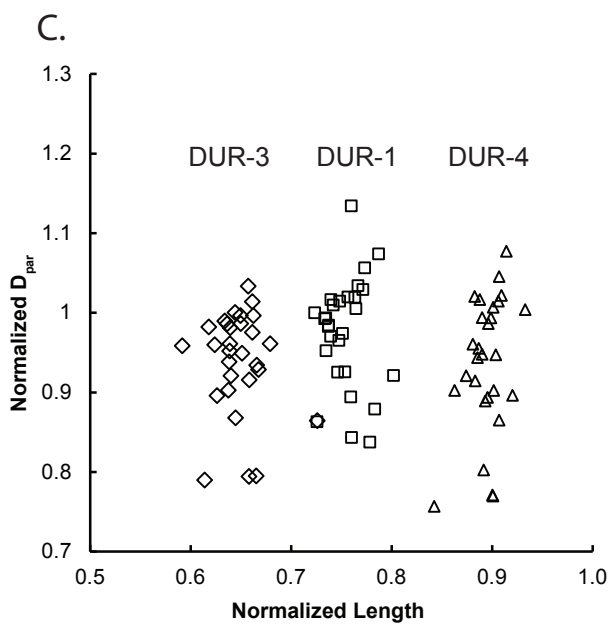
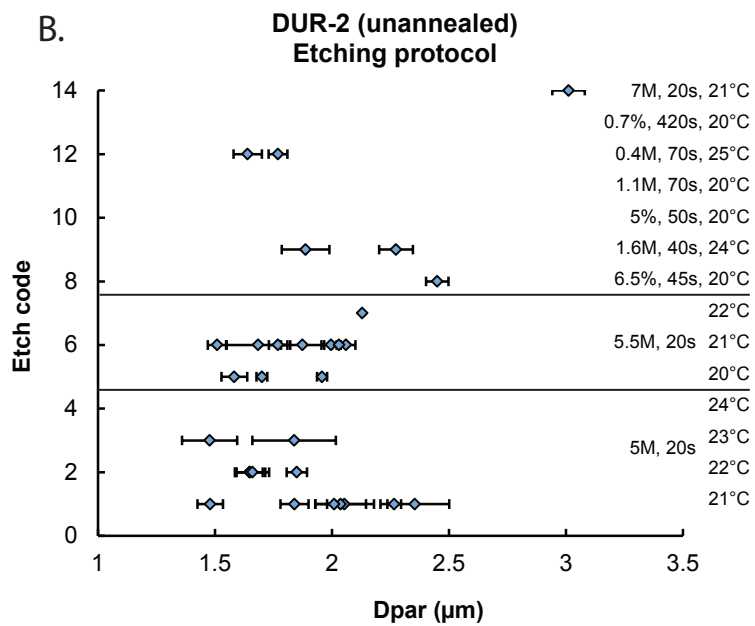
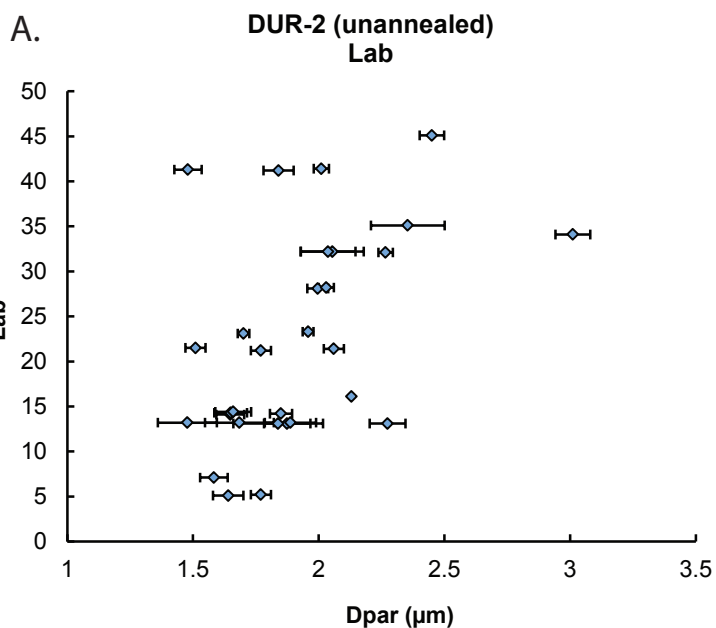












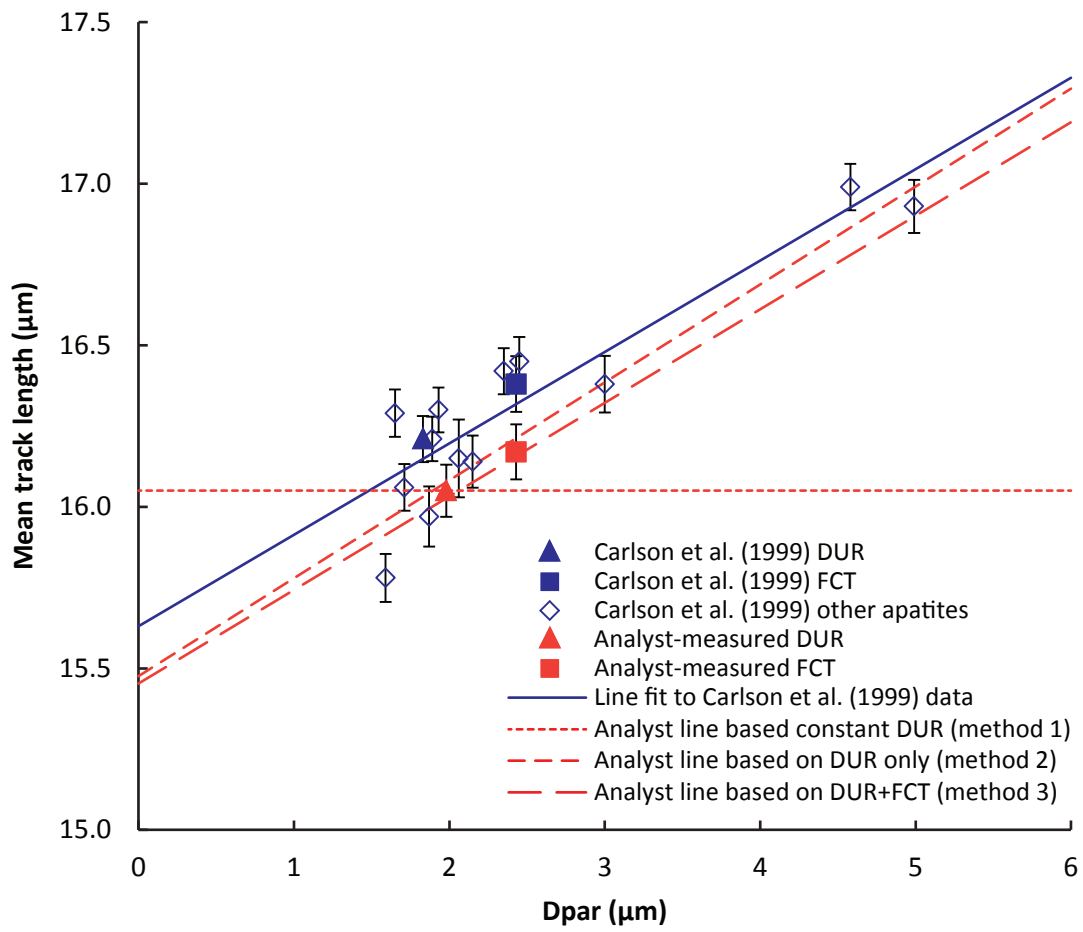


Table 1a: Survey responses for labs and analysts participating in experiment.

Lab ID	Analyst	Etch	Years	M/yr	M/yr (3 yr)	LED	Scope	Mag.	TINCLE used?	System	Standards measured
03	1	1	15.5	25	25	1	1	1600	N	2	N
05	1	12	2	20	20	1	1	1250	N	4	Internal i
	2	12	10	20	20	1	1	1250	N	4	Internal i
06	1	6	20	60	20	1	1	1250	Y	2	N
07	1	5	20	20	10	1	1	1250	?	2	N
08	1	6	3	80	80	?	2	?	N	1	N
09*	1	13	20	20	20	1	1	3840	N	3	DRi+FCi
10	1	4	14	55	35	1	1	1250	Y	2	N
12	1	?	?	?	?	?	?	?	?	?	?
13	1	6, 9, 3	24	300	300	1	1	1563	N	3	DR+FC
	2	6, 9, 3	19	300	300	1	1	1563	N	3	DR+FC
14	1	2	33	20	20	1	1	1250	N+Y	?	N
	2	2	26	2	2	1	1	1250	N+Y	?	N
	3	2	31	180	180	1	1	1250	N+Y	?	N
	4	2	15	180	180	1	1	1250	N	?	N
16	1	7	19	35	35	1	1	1250	N	2	FCs
21	A	6	0.5	2	0	1	1	1250	N	3	N
	B	6	4	15	13	1	1	1250	N	3	N
	C	6	0	0	0	1	1	1250	Y	3	N
	D	6	20	20	20	1	1	1250	N	3	N
	E	6	0	0	0	1	1	1250	Y	3	N
22	1	1	24	75	75	1	1	1600	N	3	Not recently
23	1	5	>1	>1	>1	1	1	100	N	1	DR s+i
	2	5	>1	>1	>1	1	1	100	N	1	DR s+i
	3	5	>1	>1	>1	1	1	100	N	1	DR s+i
24	1	1	>1	few	few	1	1	1000	N	3	Not suitable
25	1	1	6	37.5	37.5	1	1	1250	N	2	N
26	1	1	15	40	25	1	1	1000	N	2	N
28	1	6	17	30	30	1	1	1250	N	2	N
	2	6	5	19	9	1	1	1250	N	2	N
30	1	6	9	10	0	1	1	?	N	2	DRi Pisa
	2	6	0	0	0	2	1	?	N	2	N
32	1	1	38	100	17	1	1	1250	N	2	DR+FC

	2	1	?	?	?	?	1	1250	N	2	DR+FC
	3	1	?	?	?	?	1	1250	N	2	N
	4	1	0.01	0	0	?	1	1250	N	2	N
34	1	14	11	50	?	1	1	1250	N	2	N
35*	1	1	3	0	?	1	1	1250	N	2	N
37	1	6	0.1	0	0	1	1	1000	N	2	N
38	1	10	18	84	84	1	1	1250	N	2	DRi Pisa
39	1	6	10	20	20	1	1	1000	N	2	N
	2	6	0.1	0	0	1	1	1000	N	2	N
41	1	1	15	93	60	1	1	1250	N	2	N
	2	1	25	30	30	1	1	1250	N	2	N
	3	1	3	25	25	1	1	1250	N	2	FCs
	4	1	5	50	50	1	1	1250	N	2	DRs+FCs
	5	1	36	75	0	1	1	1600	N	2	N
	6	1	2	40	40	1	1	1600	N	2	Y
43	1	11	6	15	10	1	1	1500	N	3	DRs+DRi
	2	11	8	3	3	1	1	1500	N	3	N
45	1	8	11	40	50	3	1	1250	N	2	N
47	1	6	5	75	75	1	1	1000	N	1	DR+FC
	2	6	1	15	15	1	1	1000	N	1	DR+FC
	3	6	20	5	2	1	1	1600	N	1	DR

Question marks indicate no response provided.

Etch: Codes given in Table 1b;

Years: Number of years before experiment train in fission-track analysis;

M/yr: Estimated average number of track mounts measured per year over career;

M/yr (3 yrs): Estimated average over 3 years preceding experiment;

LED: 1=LED placed over track tips; 2=LED placed tangential to internal arc of track tips;

3=LED placed on opposite tangents of tip arcs on each end of track;

Scope: 1=air; 2=oil immersion;

Mag.: Scope magnification for measurements

System: 1=Autoscan; 2=FTStage; 3=other (custom system); 4=no computer (manual);

Standards measured: DR=Durango; FC=Fish Canyon; i=induced; s=spontaneous; Pisa=measured only during Ketcham et al. (2009) experiment; Y=yes (no details provided); N=none.

\* Labs 9 and 35 experience predominantly or exclusively with zircon rather than apatite.

Table 1b: Etch codes and protocols, as specified by respondents

Etch Code	HNO <sub>3</sub> (M)	HNO <sub>3</sub> (%)	Time (s)	Temp (°C)	Notes
1	5		20	21	
2	5		20	22	
3	5		20	23	
4	5		20	24	
5	5.5		20	20	
6	5.5		20	21	
7	5.5		20	22	
8		6.5	45	20	
9	1.6		40	24	
10		5	50	20	
11	1.1		70	20	
12	0.4	2.5	70	25	
13		0.7	Up to 420	20	Variable etch time
14	7		20	21	

Table 4: Data for sample DUR-1.

Lab	Anal.	$N_l$	$l_m$ ( $\mu\text{m}$ )	$\sigma_l$ ( $\mu\text{m}$ )	$\phi_m$ ( $^\circ$ )	$\sigma_s$ ( $^\circ$ )	$l_{c,\text{fit}}$ ( $\mu\text{m}$ )	$l_{a,\text{fit}}$ ( $\mu\text{m}$ )	$\sigma_e$ ( $\mu\text{m}$ )	$l_{c,\text{mod(B03)}}$ ( $\mu\text{m}$ )	$l_{c,\text{mod(C99)}}$ ( $\mu\text{m}$ )	$D_{\text{par}}$ ( $\mu\text{m}$ )	Notes
03	1	100	12.32(10)	0.95	--	--	--	--	--	--	--	--	
05	1	100	12.45(10)	0.96	50	19	13.59(19)	11.75(12)	0.81	13.50(06)	13.79(06)	1.86(02)	
05	2	100	12.55(16)	1.56	44	23	12.76(14)	12.34(14)	1.56	13.40(14)	13.66(13)	1.87(02)	
06	1	100	11.85(09)	0.91	54	22	12.94(20)	11.28(11)	0.77	13.09(06)	13.43(06)	--	
07	1	100	11.70(09)	0.88	59	16	13.22(29)	11.21(11)	0.76	13.11(06)	13.49(05)	1.57(04)	
08	1	221	12.19(08)	1.16	--	--	--	--	--	--	--	--	
09	1	50	12.35(14)	0.98	--	--	--	--	--	--	--	--	
10	1	100	11.94(09)	0.90	--	--	--	--	--	--	--	--	
12	1	93	12.10(10)	0.98	52	22	12.98(19)	11.55(12)	0.88	13.22(07)	13.54(07)	--	
13	1	210	12.18(06)	0.91	59	17	13.81(19)	11.60(07)	0.74	13.44(04)	13.78(03)	1.79(03)	Etch 6
13	1	201	12.45(07)	0.97	55	19	13.95(15)	11.77(08)	0.77	13.59(04)	13.90(04)	2.21(06)	Etch 9
13	1	204	12.16(07)	0.97	60	15	14.15(23)	11.54(08)	0.81	13.45(04)	13.80(04)	1.81(10)	Etch 3
13	2	200	12.60(06)	0.80	55	16	13.55(18)	12.13(09)	0.73	13.67(04)	13.98(04)	1.51(04)	Etch 6
13	2	200	12.38(07)	0.97	57	16	14.00(19)	11.76(08)	0.82	13.57(04)	13.90(04)	1.86(06)	Etch 9
13	2	200	12.18(06)	0.85	58	15	13.19(20)	11.79(08)	0.80	13.43(04)	13.77(04)	1.27(06)	Etch 3
14	1	74	12.21(12)	1.01	57	19	13.91(29)	11.51(13)	0.83	13.43(08)	13.76(07)	--	TINT
14	1	31	11.93(14)	0.75	62	13	13.43(67)	11.50(20)	0.66	13.31(09)	13.68(08)	--	TINCLE
14	1	105	12.12(09)	0.95	58	17	13.84(27)	11.49(11)	0.78	13.39(06)	13.74(05)	1.56(03)	Combined
14	2	105	12.16(07)	0.71	61	21	12.78(22)	11.92(10)	0.67	13.41(05)	13.75(05)	--	TINT
14	2	13	12.67(26)	0.90	54	22	13.23(53)	12.36(33)	0.86	13.70(18)	13.98(17)	--	TINCLE
14	2	118	12.21(07)	0.75	60	21	12.86(20)	11.96(10)	0.71	13.44(05)	13.78(05)	1.56(03)	Combined
14	3	106	12.15(09)	0.92	59	17	12.79(24)	11.90(11)	0.90	13.40(06)	13.75(06)	--	TINT
14	3	46	12.26(12)	0.80	59	19	13.07(35)	11.92(17)	0.74	13.47(08)	13.80(08)	--	TINCLE
14	3	152	12.18(07)	0.88	59	18	12.88(20)	11.90(09)	0.85	13.42(05)	13.76(05)	1.45(05)	Combined
14	4	150	11.98(06)	0.77	61	17	13.06(23)	11.62(09)	0.70	13.32(04)	13.68(04)	1.39(04)	
16	1	100	12.36(09)	0.90	56	19	13.50(22)	11.81(11)	0.78	13.50(06)	13.82(05)	2.12(00)	



41	3	100	11.96(10)	1.01	62	16	14.06(34)	11.39(10)	0.84	13.34(06)	13.71(05)	1.37(02)	
41	4	100	12.40(08)	0.82	63	11	15.07(58)	11.83(12)	0.69	13.69(05)	14.03(05)	1.86(02)	
41	5	100	11.43(10)	1.00	61	16	12.43(30)	11.12(11)	0.96	12.95(07)	13.35(06)	--	Cf
41	6	100	12.08(07)	0.72	64	11	13.26(47)	11.81(12)	0.68	13.47(05)	13.84(05)	--	
43	1	100	12.26(10)	0.98	--	--	--	--	--	--	--	--	
43	2	99	12.23(14)	1.38	--	--	--	--	--	--	--	--	
45	1	100	11.89(10)	1.00	--	--	--	--	--	--	--	2.45(02)	



Table 2: Data for sample DUR-2.

Lab	Anal.	$N_l$	$l_m$ ( $\mu\text{m}$ )	$\sigma_l$ ( $\mu\text{m}$ )	$\phi_m$ ( $^\circ$ )	$\sigma_s$ ( $^\circ$ )	$l_{c,\text{fit}}$ ( $\mu\text{m}$ )	$l_{a,\text{fit}}$ ( $\mu\text{m}$ )	$\sigma_e$ ( $\mu\text{m}$ )	$l_{c,\text{mod}(B03)}$ ( $\mu\text{m}$ )	$l_{c,\text{mod}(C99)}$ ( $\mu\text{m}$ )	$D_{\text{par}}$ ( $\mu\text{m}$ )	Notes
03	1	100	16.07(09)	0.87	--	--	--	--	--	--	--	--	
05	1	100	16.38(08)	0.82	54	20	16.78(18)	16.16(12)	0.80	16.55(06)	16.63(06)	1.64(03)	
05	2	100	16.23(09)	0.87	45	23	16.24(14)	16.21(14)	0.87	16.39(07)	16.46(07)	1.77(02)	
06	1	100	15.25(09)	0.90	56	18	15.55(21)	15.09(12)	0.89	15.68(07)	15.82(06)	--	
07	1	101	15.95(07)	0.70	54	14	16.28(24)	15.78(14)	0.69	16.22(05)	16.32(05)	1.58(03)	
08	1	211	15.34(09)	1.33	--	--	--	--	--	--	--	--	
09	1	50	16.43(11)	0.75	--	--	--	--	--	--	--	--	
10	1	100	15.89(09)	0.92	--	--	--	--	--	--	--	--	
10	2	100	15.87(08)	0.84	59	16	16.04(26)	15.80(12)	0.84	16.17(06)	16.28(06)	--	Repeat
12	1	100	15.29(12)	1.18	51	23	15.62(16)	15.06(12)	1.17	15.68(09)	15.81(09)	--	
13	1	200	16.57(06)	0.84	59	16	17.30(18)	16.30(08)	0.81	16.72(05)	16.79(04)	1.87(05)	Etch 6
13	1	205	16.84(05)	0.78	55	17	17.48(15)	16.51(09)	0.74	16.92(04)	16.97(04)	2.27(04)	Etch 9
13	1	203	16.49(07)	0.94	58	15	17.12(19)	16.24(09)	0.92	16.65(05)	16.73(05)	1.84(09)	Etch 3
13	2	200	16.59(06)	0.80	57	16	17.25(18)	16.28(09)	0.76	16.72(04)	16.79(04)	1.68(07)	Etch 6
13	2	200	16.80(07)	0.94	56	17	17.20(16)	16.60(09)	0.93	16.87(05)	16.93(05)	1.89(05)	Etch 9
13	2	200	16.77(06)	0.91	57	17	17.35(17)	16.51(09)	0.89	16.86(05)	16.92(05)	1.48(06)	Etch 3
14	1	74	15.95(10)	0.83	59	15	16.35(32)	15.79(14)	0.83	16.24(07)	16.35(07)	--	TINT
14	1	27	16.22(18)	0.89	53	17	17.13(41)	15.71(24)	0.83	16.44(14)	16.52(13)	--	TINCLE
14	1	100	16.02(09)	0.87	58	16	16.68(25)	15.74(12)	0.83	16.29(06)	16.39(06)	1.65(03)	Combined
14	2	88	15.94(08)	0.73	59	14	16.18(30)	15.84(14)	0.72	16.23(06)	16.34(05)	--	TINT
14	2	32	16.38(13)	0.74	61	17	16.46(43)	16.35(20)	0.74	16.56(10)	16.64(09)	--	TINCLE
14	2	120	16.06(07)	0.76	59	14	16.21(25)	16.00(12)	0.75	16.32(05)	16.42(05)	1.85(02)	Combined
14	3	129	15.45(11)	1.28	59	17	15.95(22)	15.25(10)	1.27	15.87(08)	16.01(07)	--	TINT
14	3	41	15.86(16)	1.00	57	18	16.75(38)	15.45(19)	0.95	16.17(12)	16.28(11)	--	TINCLE
14	3	170	15.55(09)	1.23	59	17	16.18(19)	15.29(09)	1.21	15.94(07)	16.07(06)	1.65(03)	Combined
14	4	150	15.39(08)	0.93	57	15	15.51(21)	15.33(11)	0.93	15.80(06)	15.94(05)	1.66(04)	

16	1	100	16.48(07)	0.72	60	17	16.78(24)	16.37(12)	0.71	16.64(05)	16.71(05)	2.13(00)	
20	1	94	15.53(10)	0.92	62	12	15.66(35)	15.49(13)	0.92	15.94(07)	16.08(07)	--	
21	1	232	15.93(06)	0.88	49	21	16.09(11)	15.81(09)	0.88	16.17(05)	16.26(04)	--	
21	2	256	15.56(05)	0.87	51	18	15.94(12)	15.31(08)	0.86	15.90(04)	16.02(04)	1.77(02)	
21	3	125	14.45(15)	1.63	51	20	14.32(15)	14.54(12)	1.63	15.01(12)	15.19(11)	--	
21	4	253	16.24(06)	0.88	58	20	16.42(12)	16.16(07)	0.88	16.45(04)	16.53(04)	2.06(02)	
21	5	308	14.66(05)	0.92	54	19	14.90(11)	14.52(07)	0.91	15.22(04)	15.39(04)	1.51(02)	
22	1	99	15.46(07)	0.69	60	14	15.55(29)	15.43(13)	0.69	15.87(05)	16.01(05)	--	
23	1	100	15.83(06)	0.58	63	13	15.97(34)	15.78(12)	0.58	16.16(04)	16.28(04)	1.70(01)	
23	2	99	15.82(07)	0.74	60	19	15.85(22)	15.81(11)	0.75	16.13(06)	16.24(05)	--	
23	3	100	15.78(07)	0.70	57	18	16.12(21)	15.62(12)	0.68	16.10(05)	16.21(05)	1.96(01)	
24	1	102	16.11(08)	0.79	61	14	16.35(30)	16.03(12)	0.79	16.36(06)	16.46(05)	--	
25	1	100	13.59(19)	1.85	57	16	14.76(27)	13.11(12)	1.81	14.47(13)	14.73(12)	--	
26	1	100	15.74(08)	0.76	57	17	16.18(24)	15.53(12)	0.75	16.07(06)	16.19(05)	--	
28	1	100	15.33(11)	1.13	52	19	15.42(18)	15.27(13)	1.13	15.71(09)	15.84(09)	2.00(02)	
28	2	100	14.67(16)	1.64	54	24	14.74(16)	14.62(11)	1.63	15.20(13)	15.37(12)	2.03(01)	
30	1	101	15.97(07)	0.72	62	14	16.39(30)	15.84(11)	0.71	16.27(05)	16.38(05)	--	
30	2	104	15.98(11)	1.09	64	12	15.48(33)	16.12(12)	1.09	16.28(08)	16.39(07)	--	
32	1	101	16.01(07)	0.70	61	14	15.92(29)	16.04(13)	0.69	16.28(05)	16.39(05)	--	Aliquot a
32	1	103	16.13(08)	0.81	57	18	16.51(21)	15.96(11)	0.79	16.37(06)	16.46(06)	2.27(01)	Aliquot b
32	1	104	16.52(10)	0.99	57	18	16.82(21)	16.38(12)	0.98	16.66(07)	16.73(07)	--	Aliquot c
32	2	102	15.72(09)	0.87	53	19	16.35(19)	15.33(12)	0.82	16.04(06)	16.15(06)	2.05(06)	Aliquot a
32	2	103	15.88(07)	0.75	56	19	16.01(20)	15.81(12)	0.74	16.16(06)	16.27(05)	--	Aliquot b
32	2	102	15.68(08)	0.83	55	19	16.30(20)	15.37(12)	0.79	16.02(06)	16.14(06)	2.04(05)	Aliquot c
32	3	100	15.98(09)	0.86	62	16	16.24(26)	15.89(11)	0.86	16.27(07)	16.37(06)	--	
32	4	100	15.39(11)	1.12	55	17	15.13(21)	15.54(14)	1.12	15.77(09)	15.90(09)	--	
34	1	100	15.61(10)	0.96	59	16	15.84(25)	15.53(12)	0.96	15.99(07)	16.11(07)	3.01(03)	
35	1	56	15.82(15)	1.11	53	19	16.63(27)	15.35(17)	1.05	16.13(11)	16.24(10)	2.35(07)	
37	1	50	15.88(08)	0.57	55	23	16.09(24)	15.77(16)	0.56	16.16(06)	16.26(06)	--	

38	1	100	15.88(10)	0.97	56	21	16.14(19)	15.73(12)	0.97	16.16(08)	16.26(07)	--	
39	1	80	16.45(09)	0.80	--	--	--	--	--	--	--	--	
39	2	47	15.54(16)	1.42	--	--	--	--	--	--	--	--	
41	1	100	16.00(09)	0.88	--	--	--	--	--	--	--	--	
41	2	100	16.25(07)	0.70	--	--	--	--	--	--	--	1.84(03)	
41	3	100	15.87(10)	1.01	57	15	15.99(25)	15.82(13)	1.01	16.17(08)	16.28(07)	1.48(03)	
41	4	100	16.61(07)	0.75	61	12	16.82(35)	16.54(13)	0.74	16.73(06)	16.81(05)	2.01(02)	
41	5	100	16.08(08)	0.84	59	13	16.44(30)	15.95(13)	0.83	16.34(06)	16.44(06)	--	Cf
41	6	100	15.90(10)	1.00	61	10	16.43(41)	15.73(15)	0.99	16.21(07)	16.33(07)	--	
43	1	100	15.29(14)	1.37	--	--	--	--	--	--	--	--	
43	2	99	15.27(13)	1.26	--	--	--	--	--	--	--	--	
45	1	100	16.43(09)	0.93	--	--	--	--	--	--	--	2.45(02)	
47	1	100	16.05(08)	0.81	53	22	16.70(18)	15.66(12)	0.75	16.30(06)	16.40(06)	--	
47	2	100	15.87(08)	0.84	53	22	16.53(18)	15.47(12)	0.78	16.16(07)	16.26(06)	--	
47	3	104	15.67(07)	0.75	56	18	16.05(21)	15.49(12)	0.74	16.02(05)	16.14(05)	--	

---

Table 5: Data for sample DUR-3.

Lab	Anal.	$N_l$	$l_m$ ( $\mu\text{m}$ )	$\sigma_l$ ( $\mu\text{m}$ )	$\phi_m$ ( $^\circ$ )	$\sigma_s$ ( $^\circ$ )	$N_{\text{cell}}$	$l_{c,\text{fit}}$ ( $\mu\text{m}$ )	$l_{a,\text{fit}}$ ( $\mu\text{m}$ )	$\sigma_e$ ( $\mu\text{m}$ )	$l_{c,\text{mod(B03)}}$ ( $\mu\text{m}$ )	$l_{c,\text{mod(C99)}}$ ( $\mu\text{m}$ )	$D_{\text{par}}$ ( $\mu\text{m}$ )	Notes
03	1	100	10.50(10)	0.97	--	--	--	--	--	--	--	--	--	
05	1	100	10.45(13)	1.27	48	18	99	12.38(21)	9.35(12)	0.91	12.02(07)	12.46(06)	1.48(02)	
05	2	100	10.69(15)	1.50	47	21	99	12.19(18)	9.68(12)	1.21	12.16(09)	12.56(08)	1.62(02)	
06	1	100	10.43(10)	0.99	50	16	100	11.87(21)	9.57(12)	0.78	11.98(05)	12.43(05)	--	
07	1	102	9.44(18)	1.82	56	18	90	11.57(27)	9.23(12)	0.97	11.81(06)	12.30(06)	1.52(02)	
08	1	234	10.22(09)	1.35	--	--	--	--	--	--	--	--	--	
09	1	50	11.23(16)	1.13	--	--	--	--	--	--	--	--	--	
10	1	100	10.50(12)	1.18	--	--	--	--	--	--	--	--	--	
12	1	99	9.76(14)	1.42	57	19	93	11.30(26)	9.46(12)	0.91	11.86(06)	12.37(06)	--	
13	1	204	10.60(09)	1.27	55	18	199	12.65(18)	9.87(08)	0.83	12.30(04)	12.75(04)	1.84(03)	Etch 6
13	1	204	11.14(08)	1.17	49	19	203	12.74(14)	10.21(08)	0.86	12.51(05)	12.89(04)	2.22(02)	Etch 9
13	1	200	10.56(09)	1.29	54	17	199	12.80(20)	9.65(08)	1.04	12.24(05)	12.69(05)	1.69(13)	Etch 3
13	2	200	10.78(12)	1.63	51	18	194	13.11(18)	9.82(08)	0.98	12.40(05)	12.82(05)	1.66(04)	Etch 6
13	2	200	10.94(10)	1.36	51	18	196	12.95(17)	10.01(08)	0.93	12.46(05)	12.87(04)	1.79(06)	Etch 9
13	2	175	10.68(09)	1.17	53	15	173	12.55(22)	9.91(09)	0.89	12.29(05)	12.73(04)	1.46(07)	Etch 3
14	1	118	10.65(10)	1.07	54	17	117	12.43(23)	9.89(11)	0.78	12.26(05)	12.70(05)	--	TINT
14	1	34	10.89(21)	1.19	56	19	34	12.25(38)	10.23(20)	1.05	12.40(13)	12.83(12)	--	TINCLE
14	1	152	10.70(09)	1.10	54	18	151	12.38(20)	9.97(09)	0.85	12.29(05)	12.73(05)	1.42(03)	Combined
14	2	85	10.60(13)	1.18	52	17	84	12.65(29)	9.73(13)	0.67	12.22(05)	12.66(05)	--	TINT
14	2	17	11.12(22)	0.89	50	22	17	12.26(41)	10.36(28)	0.64	12.44(11)	12.81(11)	--	TINCLE
14	2	102	10.69(11)	1.15	52	18	101	12.53(23)	9.84(12)	0.68	12.26(05)	12.69(04)	1.47(03)	Combined
14	3	119	10.14(16)	1.70	55	20	115	12.72(23)	9.29(10)	1.03	12.07(07)	12.54(06)	--	TINT
14	3	34	10.57(19)	1.09	55	20	34	12.14(41)	9.85(19)	0.94	12.19(11)	12.64(10)	--	TINCLE
14	3	153	10.24(13)	1.59	55	20	149	12.62(20)	9.41(09)	1.02	12.10(06)	12.56(05)	1.31(04)	Combined
14	4	150	9.93(11)	1.30	56	16	147	11.95(26)	9.27(09)	0.96	11.90(05)	12.41(04)	1.44(04)	
16	1	100	10.84(11)	1.14	53	20	100	13.06(24)	9.82(11)	0.74	12.38(06)	12.80(05)	--	



39	2	50	10.32(14)	1.25	--	--	--	--	--	--	--	--	--	--
41	1	101	10.34(12)	1.23	--	--	--	--	--	--	--	--	--	--
41	2	105	10.31(10)	1.02	--	--	--	--	--	--	--	--	--	1.82(03)
41	3	100	9.91(13)	1.26	59	14	98	12.63(52)	9.30(12)	0.83	11.98(05)	12.51(04)	1.42(02)	
41	4	100	10.41(11)	1.05	51	18	98	10.76(19)	10.31(14)	0.86	12.01(08)	12.45(08)	1.80(02)	
41	5	100	10.25(12)	1.17	56	16	99	12.25(32)	9.53(12)	0.95	12.07(06)	12.56(06)	--	Cf
41	6	101	10.09(09)	0.89	59	11	100	12.60(55)	9.47(13)	0.66	12.03(04)	12.54(04)	--	
43	1	96	10.68(10)	1.00	--	--	--	--	--	--	--	--	--	--
43	2	95	11.02(14)	1.33	--	--	--	--	--	--	--	--	--	--
45	1	100	10.89(10)	1.00	--	--	--	--	--	--	--	--	--	2.44(07)
47	1	100	10.29(14)	1.39	53	20	100	13.01(25)	9.09(10)	0.92	12.05(06)	12.52(05)	--	
47	2	100	10.18(14)	1.36	53	20	100	12.87(25)	8.97(10)	0.88	11.97(06)	12.44(05)	--	
47	3	111	9.18(19)	1.99	57	18	96	12.08(28)	8.88(11)	0.90	11.64(06)	12.17(05)	--	

---

Table 3: Data for sample DUR-4.

Lab	Anal.	$N_l$	$l_m$ $\mu\text{m}$	$\sigma_l$ $\mu\text{m}$	$\phi_m$ $^\circ$	$\sigma_s$ $^\circ$	$l_{c,\text{fit}}$ $\mu\text{m}$	$l_{a,\text{fit}}$ $\mu\text{m}$	$\sigma_e$ $\mu\text{m}$	$l_{c,\text{mod}(B03)}$ $\mu\text{m}$	$l_{c,\text{mod}(C99)}$ $\mu\text{m}$	$D_{\text{par}}$ $\mu\text{m}$	Notes
03	1	100	14.12(08)	0.80	--	--	--	--	--	--	--	--	
05	1	100	14.13(09)	0.88	53	17	14.88(21)	13.74(12)	0.82	14.83(06)	15.03(06)	1.48(02)	
05	2	100	14.19(12)	1.24	53	17	14.42(20)	14.06(13)	1.23	14.85(10)	15.05(09)	1.63(02)	
06	1	100	13.89(08)	0.83	60	19	14.87(25)	13.52(11)	0.76	14.71(06)	14.94(05)	--	
07	1	105	14.38(09)	0.88	56	14	15.60(30)	13.85(13)	0.81	15.05(06)	15.25(05)	1.59(02)	
08	1	175	14.36(08)	1.03	--	--	--	--	--	--	--	--	
09	1	50	14.65(08)	0.60	--	--	--	--	--	--	--	--	
10	1	100	14.46(11)	1.15	--	--	--	--	--	--	--	--	
12	1	100	13.82(10)	1.04	60	18	15.22(25)	13.32(11)	0.92	14.67(07)	14.90(06)	--	
13	1	202	14.80(06)	0.83	59	17	15.94(18)	14.37(08)	0.74	15.39(04)	15.57(04)	1.67(03)	Etch 6
13	1	203	14.88(06)	0.82	56	19	15.70(15)	14.50(08)	0.75	15.42(04)	15.59(04)	2.08(05)	Etch 9
13	1	202	14.68(06)	0.83	61	15	15.78(22)	14.34(08)	0.78	15.32(04)	15.51(04)	1.83(11)	Etch 3
13	2	200	14.86(06)	0.79	58	16	15.53(18)	14.57(08)	0.76	15.41(04)	15.58(04)	1.50(06)	Etch 6
13	2	200	15.15(05)	0.73	56	17	15.65(16)	14.90(09)	0.71	15.61(04)	15.76(04)	1.70(04)	Etch 9
13	2	200	14.96(06)	0.80	33	141	15.60(22)	14.73(09)	0.78	15.51(04)	15.68(04)	1.19(07)	Etch 3
14	1	109	14.27(07)	0.75	60	16	14.87(26)	14.04(11)	0.73	14.99(05)	15.20(05)	--	TINT
14	1	41	14.22(14)	0.88	56	14	14.45(40)	14.12(19)	0.88	14.93(11)	15.15(10)	--	TINCLE
14	1	150	14.26(06)	0.79	59	15	14.75(22)	14.07(10)	0.78	14.97(05)	15.18(04)	1.45(03)	Combined
14	2	63	14.43(08)	0.60	58	13	15.01(39)	14.19(17)	0.58	15.10(06)	15.30(05)	--	TINT
14	2	38	14.81(10)	0.63	59	15	15.28(43)	14.62(20)	0.62	15.38(07)	15.56(07)	--	TINCLE
14	2	101	14.57(06)	0.64	58	14	15.10(29)	14.36(13)	0.62	15.20(05)	15.40(04)	1.60(03)	Combined
14	3	90	13.89(09)	0.89	57	16	14.71(27)	13.53(13)	0.85	14.68(07)	14.91(06)	--	TINT
14	3	70	14.18(14)	1.13	58	19	15.74(29)	13.56(13)	0.98	14.92(09)	15.13(08)	--	TINCLE
14	3	160	14.01(08)	1.01	57	17	15.20(20)	13.52(09)	0.92	14.78(05)	15.01(05)	1.27(04)	Combined
14	4	150	13.86(07)	0.84	60	16	14.60(23)	13.61(09)	0.82	14.70(05)	14.93(04)	1.28(04)	
16	1	100	14.42(09)	0.88	57	18	15.56(25)	13.92(12)	0.79	15.08(06)	15.28(06)	--	

20	1	109	13.15(12)	1.30	61	11	13.26(36)	13.11(14)	1.30	14.20(09)	14.50(08)	--	
21	A	255	13.85(06)	1.03	56	18	14.13(13)	13.71(08)	1.02	14.62(05)	14.84(05)	--	
21	B	212	13.71(07)	0.95	52	20	14.49(13)	13.23(08)	0.88	14.47(05)	14.69(04)	--	
21	C	125	12.71(18)	1.97	48	24	11.98(12)	13.37(12)	1.92	13.60(15)	13.85(14)	--	
21	D	273	14.72(05)	0.86	54	21	15.39(11)	14.35(07)	0.81	15.28(04)	15.45(04)	--	
21	E	268	13.18(06)	0.91	58	16	13.68(16)	12.98(07)	0.90	14.16(04)	14.44(04)	--	
22	1	120	13.45(06)	0.60	58	15	13.65(24)	13.36(12)	0.60	14.35(04)	14.61(04)	--	
23	1	100	14.02(07)	0.68	60	13	15.48(39)	13.54(13)	0.59	14.82(04)	15.05(04)	1.61(01)	
23	2	99	14.21(10)	0.97	60	18	14.58(24)	14.07(11)	0.96	14.94(08)	15.16(07)	--	
23	3	100	14.14(08)	0.81	57	18	14.56(22)	13.95(11)	0.79	14.86(06)	15.07(06)	1.93(01)	
24	1	104	14.18(06)	0.64	75	30	14.75(25)	13.99(10)	0.62	14.94(04)	15.16(04)	--	
25	1	100	10.08(22)	2.19	55	18	9.72(19)	10.30(13)	2.19	12.06(14)	12.53(14)	--	
26	1	100	13.92(08)	0.79	53	18	14.63(20)	13.51(12)	0.74	14.65(06)	14.87(05)	--	
28	1	100	13.94(09)	0.91	56	19	14.57(21)	13.63(12)	0.88	14.70(07)	14.92(06)	2.04(02)	
28	2	100	13.68(14)	1.37	55	21	14.78(20)	13.14(11)	1.29	14.51(09)	14.75(08)	2.04(02)	
30	1	105	13.93(09)	0.93	66	15	14.40(33)	13.81(10)	0.93	14.79(06)	15.04(06)	--	
30	2	100	12.86(11)	1.11	61	16	13.30(27)	12.71(12)	1.10	13.96(08)	14.26(08)	--	
32	1	100	13.87(08)	0.83	59	17	14.75(25)	13.55(11)	0.77	14.69(06)	14.93(05)	2.26(03)	aliquot a
32	1	101	13.86(09)	0.86	59	15	14.59(27)	13.58(12)	0.83	14.68(06)	14.91(06)	--	aliquot b
32	1	101	13.69(10)	1.01	58	17	14.23(24)	13.46(12)	0.99	14.53(08)	14.78(07)	--	aliquot c
32	2	101	14.26(08)	0.78	54	19	15.07(20)	13.82(12)	0.71	14.92(05)	15.12(05)	2.15(05)	aliquot a
32	2	101	14.51(09)	0.88	53	19	14.87(19)	14.30(12)	0.86	15.10(07)	15.28(06)	2.24(04)	aliquot b
32	2	101	14.33(07)	0.72	55	19	15.29(21)	13.87(11)	0.61	15.00(05)	15.20(04)	2.19(05)	aliquot c
32	3	110	14.68(08)	0.88	56	18	15.47(21)	14.31(11)	0.83	15.27(06)	15.45(06)	--	
32	4	100	14.23(11)	1.13	63	20	14.15(22)	14.26(11)	1.13	14.95(09)	15.16(08)	--	
34	1	100	14.37(09)	0.92	59	18	15.32(24)	14.01(11)	0.86	15.06(07)	15.26(06)	2.70(03)	
35	1	111	13.33(14)	1.49	60	15	14.30(30)	13.00(11)	1.46	14.34(09)	14.61(08)	1.78(04)	
37	1	50	14.26(07)	0.51	52	24	14.08(22)	14.37(17)	0.50	14.85(08)	15.03(08)	--	
38	1	100	14.16(12)	1.21	50	20	14.59(17)	13.87(13)	1.19	14.80(09)	14.99(09)	--	



39	1	80	14.98(08)	0.71	--	--	--	--	--	--	--	--	
41	1	100	14.54(08)	0.76	--	--	--	--	--	--	--	--	
41	2	100	14.43(08)	0.82	--	--	--	--	--	--	--	1.87(02)	
41	2	100	14.42(08)	0.80	--	--	--	--	--	--	--	--	Cf
41	3	100	14.01(09)	0.87	63	15	15.42(35)	13.63(11)	0.80	14.84(06)	15.07(05)	1.51(02)	
41	4	100	14.73(07)	0.69	62	12	15.67(41)	14.47(13)	0.67	15.36(05)	15.55(04)	1.92(02)	
41	5	100	14.31(09)	0.86	60	15	15.26(30)	13.99(11)	0.82	15.03(06)	15.24(06)	--	Cf
41	6	101	14.34(08)	0.77	63	11	15.25(44)	14.10(13)	0.76	15.08(05)	15.30(05)	--	
43	1	100	14.28(10)	1.04	--	--	--	--	--	--	--	--	
43	2	100	14.16(13)	1.27	--	--	--	--	--	--	--	--	
45	1	100	14.85(08)	0.79	--	--	--	--	--	--	--	2.32(02)	
47	1	100	13.72(09)	0.94	59	20	14.34(21)	13.45(11)	0.90	14.55(07)	14.79(07)	--	
47	2	100	13.64(09)	0.95	59	20	14.23(21)	13.39(11)	0.92	14.49(07)	14.73(07)	--	
47	3	106	13.73(12)	1.03	61	17	14.37(25)	13.51(11)	1.01	14.60(07)	14.85(07)	--	

---