DATING THE IRON AGE I/II TRANSITION IN ISRAEL: FIRST INTERCOMPARISON RESULTS

Elisabetta Boaretto[1] • A J Timothy Jull[2] • Ayelet Gilboa[3] • Ilan Sharon[4]

ABSTRACT. Nearly a decade ago, a different chronology than the conventional absolute chronology for the early Iron Age in Israel was suggested. The new, lower chronology "transfers" Iron Age I and Iron Age IIA contexts in Israel, traditionally dated to the 11th and 10th centuries BCE, to the 10th and 9th centuries, respectively. Thus, it places the Iron I|IIA transition at about 920–900 BCE. This alternative chronology carries important implications for Israelite history, historiography, and Bible research, as well as for the chronologies of other regions around the Mediterranean. Relevant radiocarbon data sets published to date, which were measured at different sites by different laboratories, were claimed to be incompatible. Therefore, the question of agreement between laboratories and dating methods needs to be addressed at the outset of any study attempting to resolve such a tight chronological dilemma. This paper addresses results pertaining to this issue as part of a comprehensive attempt to date the early Iron Age in Israel based on many sites, employing different measuring techniques in 2 laboratories. The intercomparison results demonstrate that: a) the agreement between the 2 laboratories is well within the standard in the ^{14}C community and that no bias can be detected in either laboratory; and b) calculating the Iron I|IIa transition in 3 different ways (twice independently by the measurements obtained at the 2 labs and then by combining the dates of both) indicates that the lower chronology is the preferable one.

INTRODUCTION

Since 1996, when Finkelstein (1996) first published his suggestion to lower the beginning of Iron Age IIA in Israel from the 10th century BCE (conventional, High Chronology [HC]) to the 9th century BCE (Low Chronology [LC]), the question has not been resolved. Due to the scarcity of stratified artifacts datable on their own merit to the 11th–9th BCE range, the discussion has relied mainly on correlations of Iron Age strata with historical scenarios, based mainly on differing interpretations of the biblical text.

Regions such as Cyprus and Greece, which lack substantial anchors of absolute chronology in the early Iron ages, are directly affected by the Levantine dates. Indeed, the relevance of this dating question for diverse cultural issues regarding the Mediterranean "Dark Ages" and their aftermath has been noted by various authors (e.g. Fantalkin 2001; Gilboa and Sharon 2001, 2003; Kopcke 2002; Coldstream and Mazar 2003; Coldstream 2003; Mazar 2004). Recently, one of the main protagonists of the HC, A Mazar, conceded that Iron IIA indeed encompasses the 9th century, as suggested by the LC (e.g. Mazar 2004:30–31). Thus, the debate now concerns the beginning of Iron IIA—whether or not it should it be placed at about 1000–980 BCE, as maintained by HC adherents, or about 920 BCE, as claimed by LC adherents. This difference of approximately 60 yr is very close to the radiocarbon precision limit. Can ^{14}C dating be solicited to decide this chronological dilemma?

THE TWO PREVIOUSLY PUBLISHED RADIOCARBON DATA SETS

Tel Dor

Gilboa and Sharon (2001) published a series of 22 radiometric dates from an early Iron Age stratified sequence at the coastal site of Tel Dor (Gilboa and Sharon 2001; Sharon 2001). These samples were measured by the Weizmann Institute of Science Radiocarbon Laboratory ("Rehovot")

[1]Radiocarbon Dating Laboratory, Weizmann Institute of Science, Rehovot, Israel 76100.
 Corresponding author. Email: Elisabetta.Boaretto@weizmann.ac.il.
[2]NSF-Arizona AMS Laboratory, University of Arizona, Tucson, Arizona 85721, USA.
[3]Zinman Institute of Archaeology, University of Haifa, Mt. Carmel, Haifa, Israel 31905.
[4]Institute of Archaeology, Hebrew University, Jerusalem, Israel 91905.

employing liquid scintillation counting (LSC) and one measurement was done by Beta Analytic, Inc. using accelerator mass spectrometry (AMS). The transitions between the various early Iron Age phases at Tel Dor were calculated then using an algorithm ("transition dating") developed by Sharon (2001). The Iron I|IIA transition at Dor was determined to have occurred well within the 9th century, at around 880–850 BCE. Calculating the same transitions using the OxCal software produced consistent results with the Sharon algorithm (Gilboa and Sharon 2003: Figure 21). These results are thus compatible with the Low Chronology.

Tel Rehov

Bruins et al. (2003) published 34 ^{14}C measurements from 4 consecutive early Iron Age strata at the site of Tel Rehov in the Jordan Valley. These were analyzed by AMS and gas proportional counting (GPC) at the Center for Isotope Research at Groningen. The authors concluded that the transition between Iron Age I and IIA at Tel Rehov (between strata VII and VI) dates to ~980 BCE, about a hundred years earlier than the transition date calculated at Dor. This discrepancy was brushed away on the grounds that the Dor dates were "inconsistent" (Bruins et al. 2003:316), that the stratigraphic sequence at Dor is "complicated" (Coldstream and Mazar 2003), and/or that the dates produced at the Rehovot laboratory, at least in the 1990s, were consistently ~100 yr too low (Mazar 2004:31–34). In a later publication, Mazar presented 3 more dates from the same sequence at Rehov, also measured at Groningen (GrN-26119, GrA-12889, GrA-16848 in Table 1 of Mazar 2003), which were not included in the previous publication.

We calculated the boundaries between the dates of the Rehov strata using OxCal 3.5, considering all the Rehov (single) dates measured at Groningen. One date, GrN-26112 from L1802, was not used because its stratigraphical association is not secure enough—it cannot be assigned unambiguously to either Stratum V or VI. Initially, it was judged to belong to local Phase D-3 (Stratum VII, Iron Age I or I/II transition), and then moved to D-2 (Stratum V/VI, Iron Age IIA; Mazar, personal communication). This date is used (below) only for the purpose of intercomparison, where its exact context is unimportant. Note also that we followed Bruins et al. (2003: Table S1) in assigning locus 2425 to Stratum V. The local phasing of this silo, which produced quite a few dates analyzed in several laboratories (see below), is C-1b/a?—where C-1a is Stratum IV. As our interest is in pinpointing the date of the boundary between strata VII and VI, the distinction between strata V and IV is of lesser importance.

Running this model produces 3 measurements that display poor agreement with the rest of the sequence (misfits): GrN-26119, GrA-16848, GrN-26116. Figure 1 shows the transition between Iron Age I and IIA (the strata VII/VI boundary) after omitting these outliers. While the date proposed by Mazar et al. (980 BCE) is possible at the 68% confidence level, the distribution as a whole is much more compatible with the Low Chronology of 925 BCE. In Sharon et al. (forthcoming), we also calculated the same transition, taking into consideration Rehov dates produced at other ^{14}C laboratories (Rehovot and Tucson), with similar results. This analysis of the Rehov dates is in no way comprehensive. The point is that the Tel Rehov sequence is not nearly as unambiguous nor as decisive as has been suggested (e.g. Holden 2003).

Nevertheless, the Dor and Rehov data sets are not easily reconciled. Especially, dates from Iron Age IIA strata VI and V at Rehov (Bruins et al. 2003: Figure 2, Table S1) have significant 10th-century distributions, while at Dor this century lies well within Iron Age I.

This discrepancy could be accounted for in various ways: 1) the cultural (i.e. in this case, ceramic) horizon traditionally termed Iron IIA starts at Tel Rehov about 80 yr earlier than at Tel Dor (the sites

Boundary VII/VI w/o misfits, uncombined

68.2% probability
975BC (3.4%) 965BC
950BC (64.8%) 910BC
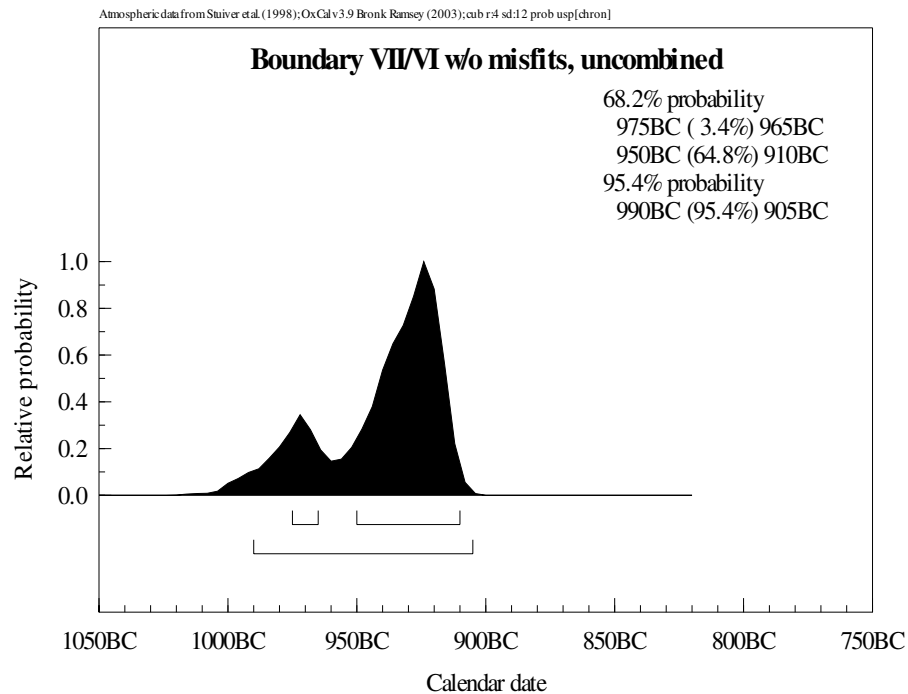95.4% probability
990BC (95.4%) 905BC

Figure 1 The Iron I/II transition at Tel Rehov based on all published Groningen dates, excluding misfits.

are about 100 km apart); 2) the terminological confusion regarding the ceramic characterization of the horizons (i.e. what the archaeologist at one site defines as Iron IIA differs from what is subsumed under this designation elsewhere); 3) the ceramic assemblages at the 2 respective sites are of very local nature and difficult to compare (this might account for either [1] and/or [2]); 4) the problem may also lie in the archaeological context from which the samples were taken (e.g. redepositions; see, e.g., Finkelstein and Piasetzky 2003); lastly, 5) the possibility of measurement bias between laboratories and/or dating methods (Mazar 2004:31–34).

Some of the archaeological questions have already been discussed (Gilboa and Sharon 2003, where the ceramic sequence of Dor is outlined). A major drawback here is that the Tel Rehov ceramic sequence has not been presented yet. In particular, the nature of the ceramic composition of the crucial Phase D-3 (= Stratum VII) is unclear. In Table S1 and the text of Bruins et al. (2003:317), it is defined as the Iron I/II transition, while in Mazar (1999:15–16) it is defined as Late Iron I. Even at this stage, however, there are enough clues to overcome these obstacles. Particularly, according to Mazar, Cypriot black-on-red and other contemporaneous Cypriot wares are evident at Tel Rehov at least as from Stratum V = Iron IIA, and possibly already in Stratum VI—the 10th century according to the Rehov team, or the last part of that century according to our calculations. At Dor, the same Cypriot wares first appear in the horizon termed there "Ir2a." Thus, these 2 horizons must be at least partially contemporary (and it is hardly conceivable that the occupations bearing these imports at inland Tel Rehov are earlier than the levels bearing the same imports at coastal Dor). However, the Ir2a horizon at Dor was ^{14}C dated to the 9th century BCE.

MATERIALS AND METHODS

Therefore, the purpose of this paper is to address the last possibility mentioned above—whether or not the problem may be found in a specific laboratory/method, and concurrently, whether ^{14}C analyses are accurate and precise enough to resolve the chronological question at hand.

With the current ^{14}C analytical precision (±30–35 yr), even ignoring additional uncertainties introduced by the calibration, differentiating between chronological hypotheses separated by less than a century is a challenge. Clearly, one needs to study as many sites as possible, using as many samples as possible. The samples must have a well-defined context and should ideally originate in occupations dating to immediately before and after the sought-for transition. It is also essential to know the precision and accuracy of the ^{14}C analyses themselves, which reflect sample preparation protocols as well as analytical procedures. It also has to be determined whether or not these factors vary during the duration of the study at any one laboratory, and most importantly, whether the accuracy of one lab is the same as that of another. Here, we address these issues as part of a larger study involving over 400 analyses from about 20 Iron Age sites in Israel. The results presented here are only those of samples that were measured by more than one laboratory or technique. The samples chosen for this intercomparison originate in archaeological contexts, the relative attribution of which, to either Iron Age I or II, are not in dispute (for most of them, see Tables 6–7 and pages 301, 372–373 of Mazar [1990]).

Laboratory Procedures

Since we were aiming for a higher resolution than ever attempted before in an archaeological project in the Levant, we imposed higher quality controls on our research than the current standard, including several methods specifically developed for this project.

To counter the effects of possible interlaboratory or intermethod bias, an elaborate protocol of replications was devised. [14]C analyses were performed in 2 laboratories, in Rehovot and at the University of Arizona ("Tucson"). In samples from which over 1 g of pure carbon could be obtained, decay counting (using LSC in Rehovot) was performed. AMS measurement was performed on all the samples in Tucson, and in almost all cases (except for extremely small samples) on triple targets. Some specimens were cleaned and pretreated in Rehovot and then analyzed in Tucson, and some were both prepared and analyzed in Tucson. For the latter, wood charcoal was homogenized in Rehovot before it was sent to Tucson, while charred seeds were sent intact.

Chemical pretreatment consisted of crushing the seeds to allow the solutions to interact with the material inside them, and charred wood was crushed and homogenized. The acid-base-acid procedure was applied in the following steps: 1 hr at 80 °C 1N HCl, 1–0.5%; NaOH as many times as required until the solution became clear; and 1 hr at 80 °C 1N HCl.

As an additional measure of control over the LSC measurements, part of the CO_2 produced for decay counting was converted to graphite and measured by AMS. Lastly, where possible, a portion of the original material from the samples was kept in the lab, so that if anomalies should be observed (in any one of the procedures), the entire experiment could be repeated.

Note that nearly all the samples were replicated in one form or another, some as many as 10 times or more. The average number of replications per sample was four.

Statistical Modeling

The first issues that need to be addressed are repeatability, reproducibility, and bias. Following Scott et al. (2003:219), we define *repeatability* as the degree of agreement between measurements on the same sample, made under identical conditions in one laboratory. *Reproducibility* refers to the degree of agreement between measurements of the same sample with different methods, under different conditions, or in different laboratories. Together, the repeatability and reproducibility constitute the precision of measurement. By *bias* (or inaccuracy), we refer to consistent differences between different methods, protocols, or laboratories. Within these issues, we need to consider the treatment of obviously aberrant results, as well as smaller deviations that may cause the actual spread of replicated results to be wider than the distribution expected under the quoted measurement error.

The Fourth International Radiocarbon Intercomparison (FIRI; Scott et al. 2003, see especially pages 213–218, 252–260, and Figures 7.1–7.7) demonstrates that while little or no bias can usually be detected between laboratories, ^{14}C measurements generally do not conform to the "ideal" standard distribution. The actual distribution under extensive replication is wider than the quoted errors. This effect is more noticeable in LSC than in AMS; it is also not related to sample age. Even after the exclusion of the more-obvious outliers and the recalculation of central moments, the observed distributions are wider than the theoretical expectation. Indeed, the case could hardly be different, for the quoted error denotes, at best, the variability within the same measurement run. Additional sources of error—such as variability due to different machines, different measurement protocols, different chemical cleaning and pretreatment, and even differences in storage conditions and length of storage prior to measurement—would all inevitably lead to a wider distribution than indicated by the individual measurement error.

Once outliers—or distributions in excess of the expected statistical errors—have been identified, the problem of how to treat them remains. The dilemma here is that, on the one hand, basing conclusions on data known to be partially faulty would not be prudent. On the other hand, the greatest danger in this sort of study is selectivity (explicitly or implicitly removing data that are not consistent with preconceived interpretations).

The experimental design we devised allows us to distinguish several types of outliers:

a) Intra-lab errors. A divergent result within a series of measurements, made under identical conditions in one laboratory. The probable source of error is the measurement itself.
b) Inter-lab inconsistencies. Divergent results between measurement methods or laboratories. The probable source of error here are laboratory procedures. In the case of decay-counting, differentiating between type (a) and type (b) is usually not possible because there rarely is enough raw material for replication.
c) Archaeological errors. Results which replicate well but radically deviate from their supposed archaeological placement, i.e. diverge significantly from other measurements of the same typo-stratigraphic horizon, or absolutely do not fit the typo-stratigraphic sequence. The probable source of error is archaeological.

Different strategies of dealing with divergent data are found in the literature. The strategy used in the current study is as follows.

First, suspect measurements were rechecked to see if some independent cause for the aberrant reading (physical or archaeological) could be located. If so, all the measurements in which the same problem was found were eliminated, whether or not they agreed with other results.

Visual and microscopic examination of some of the graphite targets that were used in the accelerator and gave divergent results had indeed shown an anomalous behavior of the graphite in the cathode during the sputtering process. The possible reason for this phenomenon is improper pressing of the graphite. All these targets were excluded from the study.

The samples that were excluded on archaeological grounds were, e.g., the Hazor XII/XI samples (RTT 3700, 3701, 3702, 3703, 3704), which produced dates in the early 2nd millennium. (But the Hazor dates are still useful for the purpose of intercomparison, residuality notwithstanding.)

As already noted above, one method used to cross-check the samples prepared for LSC in Rehovot was to extract a small amount of CO_2 during sample preparation and measure it with AMS. It turned out that these samples fluctuated widely and without any correlation with the date obtained for the same sample by AMS or LSC. We do not yet know the exact reason for this. Since this exercise is not part of the normal operational run in the laboratory, it was decided to defer investigation of this phenomenon and to simply not use results obtained this way.

An interesting and fairly consistent offset was noted on some samples prepared and dated in the Tucson laboratory, labeled "b." These samples were run differently than the rest of the Tucson targets (labeled "a" and "aa"). Before graphitization, the CO_2 from these "b" samples was run through a mass spectrometer to measure the $\delta^{13}C$. We suspect that the sample gas might have been mixed with the reference gas before the recollection for the graphitization. We have investigated this problem further and identified possible differences in procedure used by mass spectrometer operators, which might account for such differences. All the "b" targets were excluded from the analysis.

In a few cases, we noted that a single AMS batch had several outliers. This raised the possibility of either instability in the accelerator during that run or problems with the measurement of the standards. In these cases, we prepared and re-ran several of the samples included in these batches (from the material left at Rehovot). These suspicions were unfounded, and most of the re-runs faithfully replicated the previous results. Indeed, the fact that independently pretreated samples run after a 2-yr interval gave very similar results greatly enhances our confidence in the stability of the system and the AMS measurements. Such re-runs account for samples with many replications (typically 3 measurements in Tucson and 6 in Rehovot).

If, even after all these steps, a replicated series did not pass the χ^2 test, we left the potentially outlying measurements in, but used the sampled standard deviation instead of the combined error term whenever dates were combined (Bevington and Robinson 1992).

The next step of analysis is to construct a model for the collation of all of the dates together with the order of the contexts from which they were collected (as determined by stratigraphic and typological analyses). In the present study, we employ the standard Bayesian inference approach (Bronk Ramsey 1994, 1995).

THE ASSESSMENT OF BIAS

The following analyses were carried out:

1. Measurements intended to assess the effect of pretreatments: intercomparison of AMS dates of samples prepared in Rehovot and prepared in Tucson, both analyzed in Tucson.

Twenty-two samples were split between Rehovot and Tucson, pretreated in these 2 labs, and measured by AMS in Tucson, producing 90 individual measurements (Table 1). This exercise was carried out during a 3-yr interval, using different machines during normal operation time. Table 1 includes all outliers except ones that may be omitted due to some procedural error as explained above.

Table 1 Single measurements of the 22 samples divided between Tucson and Rehovot.

				Tucson		Rehovot	
Material	Site	Strata	Period	AA/T sample #[a]	^{14}C age ±1 σ (BP)	RTT sample#[b]	^{14}C age ±1 σ (BP)
Charcoal	Hazor	XII/XI	Iron I	AA40992	2945 ± 50	3700*	2975 ± 35
Charcoal	Hazor	XII/XI	Iron I	AA40993	2965 ± 50	3701*	2940 ± 30
Olive pits	Hazor	XII/XI	Iron I	AA40994	3060 ± 50	3702*	3060 ± 30
Charcoal	Hazor	XII/XI	Iron I	AA40995	3650 ± 50	3703*	3570 ± 53
Charcoal	Hazor	XII/XI	Iron I	AA40996	3370 ± 60	3704*	3375 ± 30
Olive pits	Yoqne'am	XVII (b?)	Late Iron I	T18150a	2778 ± 42	3778.3	2781 ± 45
				T18150aa	2866 ± 45	3778.4	2866 ± 45
				—	—	3778.5	2802 ± 46
				Average	2818 ± 29	*Average*	2817 ± 26
Charcoal	Yoqne'am	XIVb	Early Iron IIA	T18151a	2739 ± 35	3780.3	2635 ± 45
				T18151Aaa	2726 ± 43	3780.4	2657 ± 55
				—	—	3780.5	2663 ± 69
				Average	2739 ± 35	*Average*	2649 ± 31
Charcoal	Hazor	Xa	Iron IIA	T18152a	2731 ± 37	3783.3	2680 ± 48
				T18152aa	2674 ± 35	3783.4	2822 ± 40
				—	—	3783.5	2795 ± 38
				Average	2700 ± 27	*Average*	2777 ± 24
Olive pits	Hazor	IXa	Late IIA	T18153a	2693 ± 35	3785.4	2675 ± 48
				T18153aa	2705 ± 52	3785.5	2707 ± 46
				—	—	3785.6	2683 ± 46
				Average	2697 ± 24	*Average*	2689 ± 27
Olive pits	Hazor	Xb	Early Iron IIA	T18154a	2576 ± 66	3786.3	2615 ± 81
				T18154aa	2656 ± 35	3786.4	2448 ± 48
				—	—	3786.5	2695 ± 47
				Average	2639 ± 31	*Average*	2585 ± 126
Seeds	H. Rosh Zayit	IIa	Iron IIA	T18155a	2687 ± 35	3798.3	2763 ± 37
				T18155aa	2693 ± 47	3798.4	2753 ± 37
				—	—	3798.5	2749 ± 38
				Average	2689 ± 28	*Average*	2755 ± 22
Seeds	H. Rosh Zayit	IIa	Iron IIA	T18156a	2683 ± 35	3799.3	2729 ± 37
				T18156aa	2728 ± 68	—	—
				Average	2692 ± 31	*Average*	2729 ± 37
Charcoal	Tell Keisan	9a	Late Iron I	T18157a	2938 ± 35	3803.3	2817 ± 48
				T18157aa	2862 ± 65	3803.4	2940 ± 50
				—	—	3803.5	2997 ± 35
				—	—	3803.6	2871 ± 42
				—	—	3803.7	2798 ± 36
				—	—	3803.8	2799 ± 36
				Average	2921 ± 31	*Average*	2870 ± 82
Charcoal	Tell Keisan	13	LB | Iron I	T18158a	3022 ± 47	3804.3	2960 ± 35
				T18158aa	2985 ± 36	3804.4	2997 ± 35
				—	—	3804.5	2996 ± 35
				Average	2999 ± 29	*Average*	2984 ± 20
Olive pits	Tel Rehov	E-1b(=V)	Iron IIA	T18159a	2671 ± 35	3808.3	2693 ± 35
				T18159aa	2700 ± 37	3808.4	2671 ± 35
				—	—	3808.5	2669 ± 35
				Average	2685 ± 25	*Average*	2678 ± 20

Table 1 Single measurements of the 22 samples divided between Tucson and Rehovot. *(Continued)*

Material	Site	Strata	Period	Tucson		Rehovot	
				AA/T sample #[a]	^{14}C age $\pm 1\,\sigma$ (BP)	RTT sample #[b]	^{14}C age $\pm 1\,\sigma$ (BP)
Olive pits	Tel Rehov	D-4 (=VII)	Late? Iron I	T18150a	2889 ± 35	3809.4	2830 ± 34
				T18150aa	2952 ± 44	3809.5	2861 ± 36
				Average	2913 ± 27	*Average*	2845 ± 25
Seeds	Tell Qasile	X	Late Iron I	T18161a	2780 ± 35	3932.3	2746 ± 50
				T18161aa	2862 ± 38	3932.4	2763 ± 76
				—	—	3932.5	2683 ± 49
				—	—	3932.6	2651 ± 37
				Average	2818 ± 26	*Average*	2692 ± 24
Seeds	Megiddo	K4=VIA	Late Iron I	T18163a	2864 ± 40	3944.3	2974 ± 63
				—	—	3944.4	2982 ± 47
				—	—	3944.5	2904 ± 58
				Average	2864 ± 40	*Average*	2957 ± 31
Seeds	Megiddo	H5	Early Iron IIA	T18167a	2788 ± 38	3949.3	2821 ± 47
				T18167aa	2807 ± 42	3949.4	2899 ± 48
				Average	2796 ± 28	*Average*	2859 ± 34
Olive pits	Tel Miqne-'Ekron	IV	Late Iron I	T18168a	2872 ± 36	4282.3	2894 ± 53
				T18168a	2895 ± 38	4282.4	2837 ± 42
				—	—	4282.5	2899 ± 47
				Average	2883 ± 26	*Average*	2872 ± 27
Olive pits	Megiddo	K6 (=VIIA?)	LB | Iron I	T18169a	2907 ± 37	4499.3–1.1	2884 ± 37
				T18169aa	2876 ± 40	4499.3–1.2	2866 ± 43
				—	—	4499.3–1.3	2927 ± 39
				Average	2893 ± 27	*Average*	2894 ± 23
Olive pits	Megiddo	K6 (=VIIA?)	LB | Iron I	T18170a	3018 ± 56	4500.3.1–1	2940 ± 38
				T18170aa	2947 ± 36	4500.3.1–2	2906 ± 37
				—	—	4500.3.1–3	2909 ± 37
				Average	2968 ± 30	*Average*	2918 ± 22

[a] AA or T = pretreated and measured in Tucson by AMS.
[b] RTT = pretreated and prepared as graphite in Rehovot and measured by AMS in Tucson; * = the average of 3 measurements.

To exclude the possibility of a meaningful difference between the 2 laboratories, a Bland-Altman plot (Bland and Altman 1986) of the data from Table 1 is shown in Figure 2. In Figure 3, the same data with the analytical error are compared. In Figure 2, the *x* axis denotes the average of the Tucson and Rehovot (AMS) ^{14}C determinations for the same sample. The *y* axis denotes the difference between the Tucson average and Rehovot (AMS) average for the same sample. The mean difference and standard deviation of the differences between the 2 laboratories is 9.96 ± 57.56 yr. The mean difference value is represented in the plot by the solid line. The dashed lines represent the mean difference ±1.96*1 σ, which is indicated in the plot with the dashed line at +123 yr and –103 yr. All the points are included in the region between the 2 lines. The paired Student's *t* test provided a *t* value of 0.81 and a *p* value of 0.43. Thus, the null assumption of no difference between the labs cannot be negated at a reasonable statistical confidence level. In Figure 3, the *y* axis denotes the average of the multiple measurements of the sample prepared in Rehovot, in ^{14}C BP; the error bars are ±1 standard deviation. The *x* axis denotes the same for samples prepared in Tucson. The line of equality, *x* = *y* line, which represents the perfect match, is indicated (dashed). Approximately the same number of points fall on either side of this line, meaning that in about half of the cases, Rehovot produced higher dates than Tucson and vice versa. Also shown is the best-fit regression line, which is conditioned to pass through the origin of the axes, and its correlation coefficient. Likewise, the $\delta^{13}C$ was found to be in good agreement between the 2 laboratories.
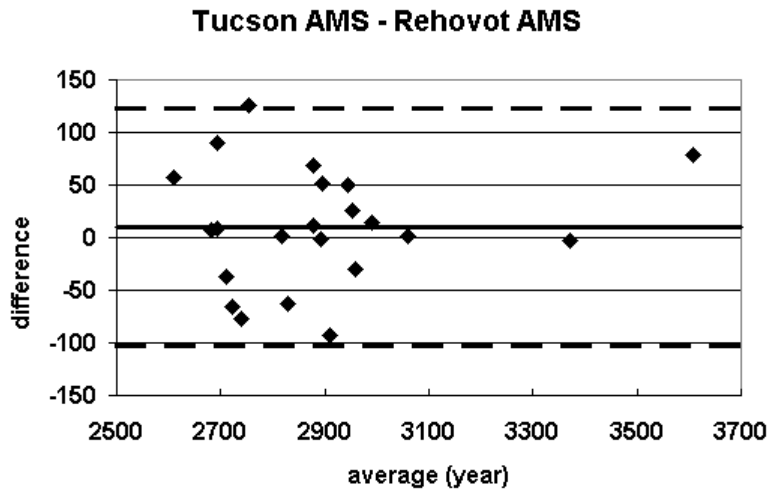
Tucson AMS - Rehovot AMS

Figure 2 Tucson AMS/Rehovot AMS pretreatment intercomparison results. The *x* axis represents the average of the measurements for the sample prepared and measured in Tucson and for the sample prepared in Rehovot and measured in Tucson; the *y* axis represents the difference of the samples prepared and measured in Tucson and the samples prepared in Rehovot and measured in Tucson. The $y = 9.96$ line (solid) and the 1.96*(±57.56) lines (dashed) are indicated.
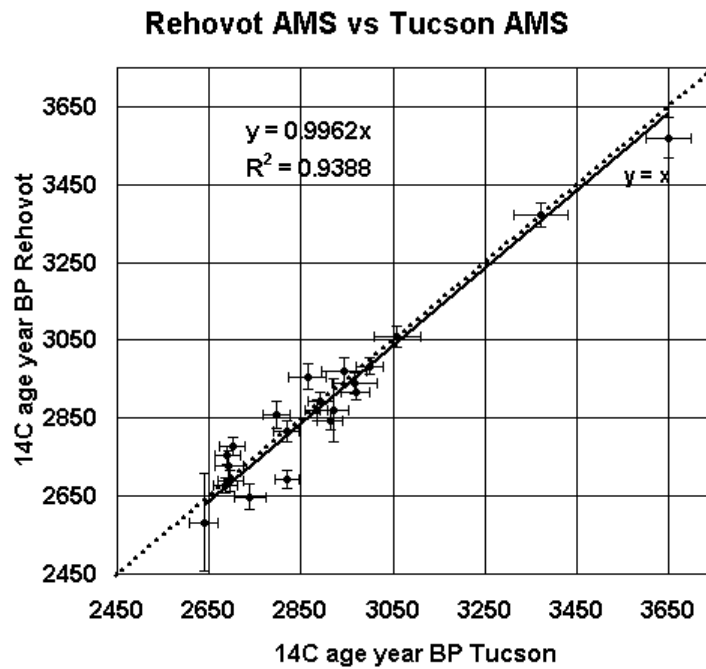
Rehovot AMS vs Tucson AMS

Figure 3 Tucson/Rehovot pretreatment intercomparison results. The *x* axis represents the samples prepared and measured in Tucson; the *y* axis represents the samples prepared in Rehovot and measured in Tucson. The $y = x$ line (dashed) and the best-fit line (solid) for the set of points are indicated.

2. Measurements intended to compare AMS and LSC dates: same pretreatment and different measurement methods.

This exercise is intended to assess the effects of dating techniques. It includes 14 samples pretreated at Rehovot and measured by LSC in Rehovot and by AMS in Tucson. Of these, 5 samples underwent 3 different procedures: LSC (pretreated and measured in Rehovot), AMS prepared at Rehovot and measured in Tucson, AMS prepared and measured in Tucson.

Table 2 LSC and AMS results for sample measurements from Rehovot and Tucson.

Material	Site	Stratum	Period	LSC sample #	LSC[a] 14C age ±1 σ (BP)	RTT sample #	AMS[b] 14C age ±1 σ (BP)	Tucson sample #	AMS[c] 14C age ±1 σ (BP)
Olive pits	Yoqne'am	XVII (b?)	Late Iron I	—	—	3777.3	2830 ± 60	—	—
				—	—	3777.4	2910 ± 45	—	—
				—	—	3777.5	2790 ± 80	—	—
				3777.1	2866 ± 25	—	2866 ± 33*	—	—
Olive pits	Yoqne'am	XVII (b?)	Late Iron I	—	—	3778.3	2781 ± 45	T18150a	2778 ± 42
				—	—	3778.4	2866 ± 45	T18150aa	2853 ± 39
				—	—	3778.5	2802 ± 46	—	—
				3778.1	2776 ± 25	—	2817 ± 26*	—	2818 ± 29*
Charcoal	Yoqne'am	XVIIa	Late Iron I	—	—	3779.3	2815 ± 45	—	—
				—	—	3779.4	2870 ± 60	—	—
				—	—	3779.5	2800 ± 55	—	—
				3779.1	2926 ± 30	—	2824 ± 30*	—	—
Charcoal	Yoqne'am	XIVb	Early Iron IIA	—	—	3780.3	2635 ± 45	T18151a	2739 ± 35
				—	—	3780.4	2657 ± 55	T18151aa	2726 ± 43
				—	—	3780.5	2663 ± 69	—	—
				3780.1	2725 ± 25	—	2649 ± 31*	—	2739 ± 35*
Seeds	H. Rosh Zayit	IIa	Iron IIA	—	—	3798.3	2763 ± 37	T18155a	2687 ± 35
				—	—	3798.4	2753 ± 37	T18155aaA	2693 ± 47
				—	—	3798.5	2749 ± 38	—	—
				3798.1	2745 ± 30	—	2755 ± 23*	—	2689 ± 27*
Seeds	H. Rosh Zayit	IIa	Iron IIA	—	—	3799.3	2729 ± 37	T158156a	2683 ± 35
				—	—	—	—	T18156aa	2728 ± 68
				3799.1	2745 ± 30	—	2729 ± 37*	—	2692 ± 31*
Charcoal	Tell Keisan	9a-b	Late Iron I	—	—	3802.3	2820 ± 50	—	—
				—	—	3802.4	2820 ± 50	—	—
				—	—	3802.5	2885 ± 50	—	—
				3802.1	2870 ± 35	—	2842 ± 29*	—	—
Charcoal	Tell Keisan	9a	Late Iron I	—	—	3803.3	2817 ± 48	T18157a	2938 ± 35
				—	—	3803.4	2940 ± 50	T18157aa	2862 ± 65
				—	—	3803.5	2997 ± 35	—	—
				—	—	3803.6	2871 ± 42	—	—
				—	—	3803.7	2798 ± 36	—	—
				—	—	3803.8	2799 ± 36	—	—
				3803.1	2893 ± 50	—	2870 ± 82*	—	2921 ± 31*
Olive pits	Tel Rehov	D-3? D-2?	Iron IIA	—	—	3807.3	2763 ± 35	—	—
				—	—	3807.4	2716 ± 35	—	—
				—	—	3807.5	2757 ± 20*	—	—
				3807.1	2760 ± 40	—	—	—	—
Seeds	Tell Qasile	X	Late Iron I	—	—	3931.3	2820 ± 55	—	—
				—	—	3931.4	2935 ± 41	—	—
				—	—	3931.5	2936 ± 41	—	—
				3931.1	2853 ± 25	—	2911 ± 26*	—	—

Table 2 LSC and AMS results for sample measurements from Rehovot and Tucson. *(Continued)*

Material	Site	Stratum	Period	LSC sample #	LSC[a] 14C age ±1 σ (BP)	RTT sample #	AMS[b] 14C age ±1 σ (BP)	Tucson sample #	AMS[c] 14C age ±1 σ (BP)
Seeds	Bet Shemesh	3	Iron IIB	—	—	3937.3	2524 ± 36	—	—
				—	—	3937.4	2427 ± 35	—	—
				—	—	3937.5	2478 ± 34	—	—
				3937.1	2500 ± 40	—	2475 ± 20*	—	—
Charcoal	Bet Shemesh	Early 3	Iron II A/B	—	—	3987.3	2844 ± 34	—	—
				—	—	3987.4	2889 ± 35	—	—
				—	—	3987.5	2952 ± 51	—	—
				3987.1	2879 ± 40	—	2882 ± 22*	—	—
Charcoal	Tel Hevron	VII	Iron I	—	—	4148.3	3013 ± 36	—	—
				—	—	4148.4	2909 ± 53	—	—
				—	—	4148.5	3025 ± 37	—	—
				4148.1	3010 ± 35	—	2998 ± 23*	—	—
Seeds	Bethsaida	VI	Iron I or II	—	—	4281.3	2775 ± 50	—	—
				—	—	4281.4	2800 ± 40	—	—
				—	—	4281.5	2780 ± 40	—	—
				4281.1	2820 ± 35	—	2786 ± 25*	—	—

[a]Samples pretreated and measured by LSC in Rehovot.
[b]Samples pretreated at Rehovot and measured by AMS in Tucson; average measurements denoted by *.
[c]Samples pretreated and measured by AMS in Tucson; average measurements denoted by *.

The LSC/AMS results are shown in Table 2 and are represented in Figures 4 and 5 (as Bland-Altman plots) and in Figure 6. Figure 4 is the Bland-Altman plot for the LSC Rehovot and AMS Rehovot results. In Figure 4, the *x* axis denotes the average of Rehovot LSC and AMS ^{14}C determinations for the same sample. The *y* axis denotes the difference between the Rehovot LSC and Rehovot (AMS) average for the same sample. The mean difference and standard deviation of the differences between the 2 methods is 15.9 ± 39.0 yr. The mean difference value is represented in the plot by the solid line; the dashed lines represent the mean average ±1.96*1 σ (indicated in the plot with the dashed line at +92.4 yr and –60.6 yr). In Figure 5, the *x* axis denotes the average of Rehovot LSC and Tucson ^{14}C determinations for the same 5 samples. The *y* axis denotes the difference between the Rehovot LSC and Tucson average for the same sample. The mean difference and standard deviation of the differences between the 2 laboratories is –5.0 ± 46.0 yr. The mean difference value is represented in the plot by the solid line. The dashed lines represent the mean average ±1.96*1 σ (indicated in the plot with the dashed line at +85.7 yr and –95.7 yr).

In Figure 6, the *y* axis for each point denotes the average of the multiple AMS measurements for samples prepared in Rehovot and measured in Tucson, in ^{14}C BP; the error bars are ±1 standard deviation. The *x* axis denotes the ^{14}C age for the same samples measured by LSC in Rehovot. The line of equality, *x = y* line, is indicated. Also shown is the best-fit regression line, which is conditioned to pass through the origin of the axes, and its correlation coefficient. With a mean average of the differences between the 2 laboratories of 15.9 yr, the paired Student's *t* test provided a *t* value of 1.45 and a *p* value of 0.17. Thus, the null assumption of no difference between the labs cannot be negated at a reasonable statistical confidence level.

Figure 6 also presents the 5 samples for which all 3 procedures were performed (the samples prepared at Tucson are identified by a cross symbol, while the same samples prepared in Rehovot are designated by a diamond). In these cases as well, it is impossible to detect a systematic difference between the laboratories that is larger than the errors quoted by them.
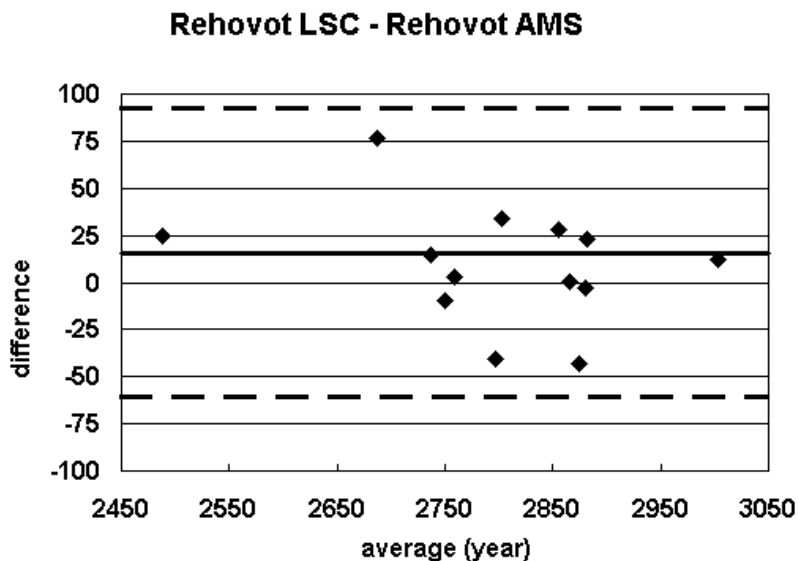
Rehovot LSC - Rehovot AMS

Figure 4 Rehovot LSC/AMS Rehovot intercomparison results. The *x* axis represents the average of the measurements for the sample prepared and measured in Rehovot by LSC and for the sample prepared in Rehovot and measured in Tucson by AMS; the *y* axis represents the difference of the samples prepared and measured in Rehovot by LSC and the samples prepared in Rehovot and measured in Tucson. The $y = 15.9$ yr line (solid) and the 1.96*(±39.0) lines (dashed) are indicated.
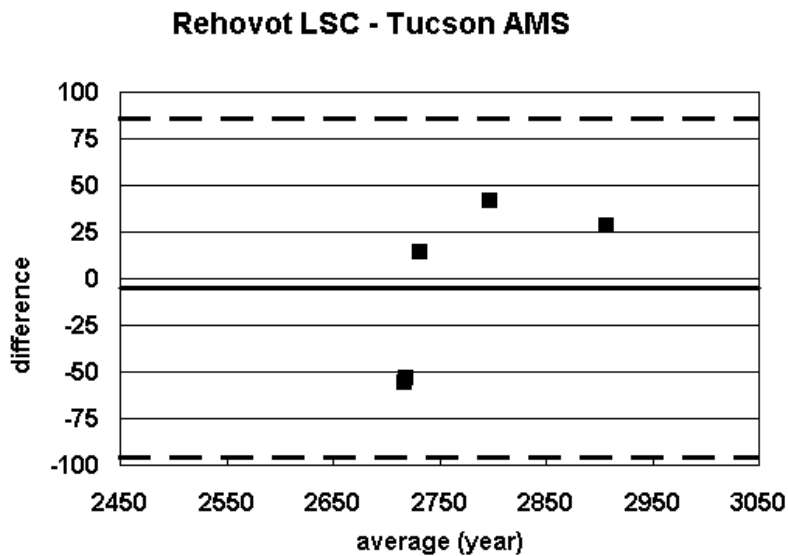
Rehovot LSC - Tucson AMS

Figure 5 Rehovot LSC/AMS Tucson intercomparison results. The *x* axis represents the average of the measurements for the sample prepared and measured in Rehovot by LSC and for the sample prepared and measured in Tucson by AMS; the *y* axis represents the difference of the samples prepared and measured in Rehovot by LSC and the samples prepared and measured in Tucson. The $y = -5.0$ yr line (solid) and the 1.96*(±46.0) lines (dashed) are indicated.
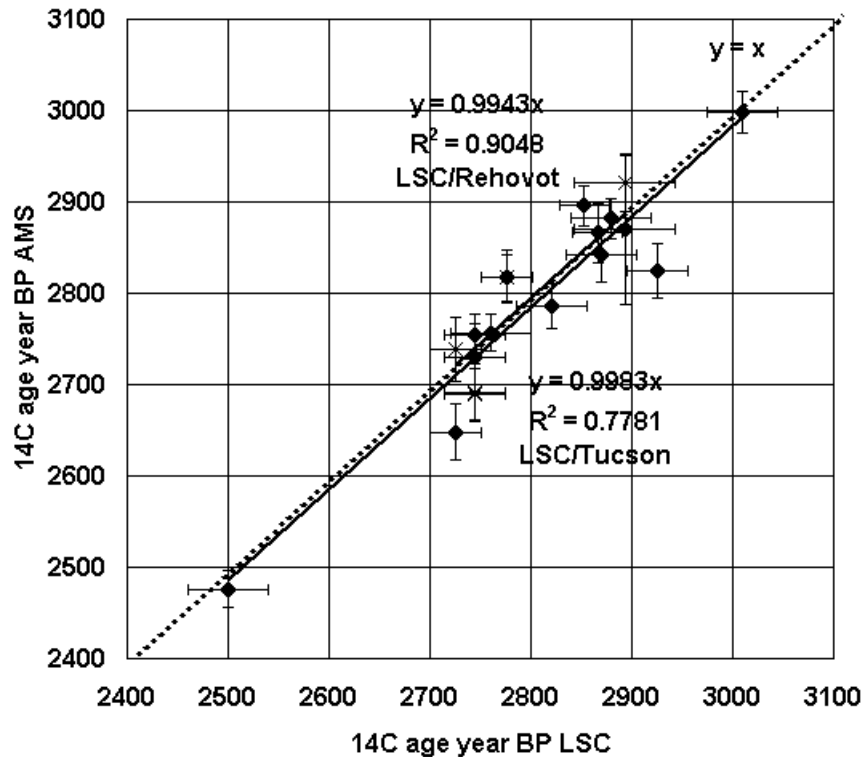
Rehovot LSC vs AMS

Figure 6 Rehovot LSC/Rehovot AMS/Tucson AMS intercomparison results. The *x* axis represents the samples prepared and measured at Rehovot by LSC; the *y* axis diamond symbols represent samples prepared in Rehovot and measured in Tucson by AMS; the *y* axis cross symbols are samples prepared and measured at Tucson by AMS. The *y* = *x* line (dashed) and the best-fit (continuous) line are indicated.

3. Intercomparison with Tel Rehov published dates measured at Groningen.

In 3 cases, dates from Tel Rehov, produced at Rehovot and/or Tucson, may be compared with dates produced from the same contexts at Groningen (published in Bruins et al. 2003; Mazar 2004). However, it must be emphasized that contrary to the results presented in the previous paragraphs, the comparison here is of more indirect nature. In each case, the material analyzed in the different laboratories is from the same locus but might not be exactly the same (a similar exercise was presented in Mazar 2004: Tables 1–2).

In averaging the different ^{14}C dates from the same locus, all sets pass the χ^2 test, even the set related to L2862, where the 2 Groningen measurements differ considerably. This indicates that, in general, there is good agreement between the laboratories. Averaging the sets of data separately for each laboratory, only in one case out of three (in the sample from Locus 1802) are the Rehovot averages separated by more than 3 σ from the Groningen (single) measurement. In the other 2 cases, the averages are the same, based on the quoted errors. The restricted number of measurements does not permit a more detailed analysis, but they definitely are in agreement and no bias can be detected.

Table 3 Groningen, Rehovot, and Tucson data for Tel Rehov samples taken from the same loci.[a]

Material	Area-phase (stratum)	Locus	Groningen		Rehovot		Tucson	
			Sample #	^{14}C age ±1 σ (BP)	Sample #	^{14}C age ±1 σ (BP)	Sample #	^{14}C age ±1 σ (BP)
Olive pits	D-3	2862	GrA-19033[b]	2835 ± 45	RTT-3805	2775 ± 35	—	—
		—	GrN-26119[c]	2720 ± 30	—	2810 ± 35	—	—
		—	—	—	—	2815 ± 35[d]	—	—
Olive pits	D-2 (V)	1802	GrN-26112	2805 ± 15	RTT-3807	2765 ± 35	—	—
		—	—	—	—	2715 ± 35	—	—
		—	—	—	RT-3807	2795 ± 40	—	—
		—	—	—	—	2760 ± 40	—	—
		—	—	—	—	2785 ± 35	—	—
Olive pits	D-4 (VII)	1836	GrN-26121	2890 ± 30	RTT-3809	2830 ± 35	T18150A	2890 ± 35
		—	GrA-18825	2870 ± 50	—	2860 ± 35	—	2950 ± 45

[a]GrN = Groningen decay counting; GrA = Groningen AMS; RTT = Rehovot AMS; RT = Rehovot LSC; T = Tucson AMS.
[b]In Table 1 of Mazar (2004), this sample is attributed to L2862, and in Bruins et al. (2003) to L4815, both of Phase D-3.
[c]From Table 1 of Mazar (2004). The obviously outlying result may have been due to the fact that the sample was small and was diluted to produce enough material for GPC (van der Plicht, personal communication).
[d]Cited erroneously as 2785 ± 25 in Mazar (2004). In the errata addendum to the paper, it is corrected to 2800 ± 20, which is the average of these 3 measurements.
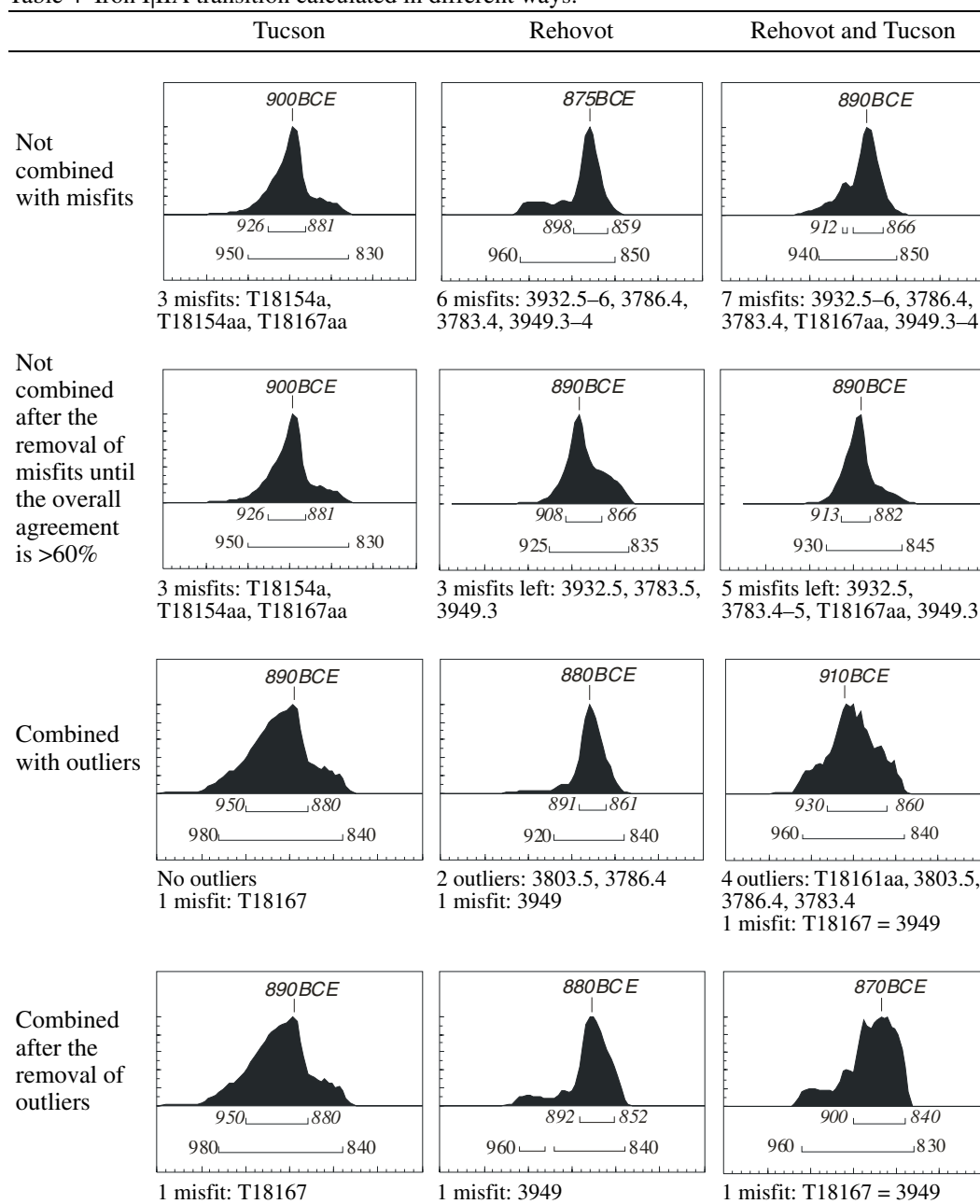
THE IRON I/IIA TRANSITION DATE USING THE INTERCOMPARISON EXERCISE DATA

The archaeological contexts which yielded the samples presented here were grouped into 2 broad archaeological horizons for the purposes of this exercise—those originating from Iron Age I contexts and those of Iron Age IIA (see Table 1). The simplest model, | Iron Age I | Iron Age IIA |, was used with the "|" standing for a "boundary" term, and the object-of-interest being the temporal distribution of the middle boundary (Bronk Ramsey 2000). This has the effect of reducing the archaeological variables in the system, *inter alia*, meaning that the intermediate horizon, which we termed "Ir1|2" (Gilboa and Sharon 2003: Table 21), is not represented in this paper. This omission, or possible gap, does not affect the results of intercomparison, but needs be taken into account when interpreting the results of the models. Within this general model, several different sub-models were tested:

a) Individual measurements from each laboratory, and both laboratories together, no combination.
b) As above, but results which were flagged as being in "poor agreement," with the model being removed stepwise, starting with the poorest, until the general agreement index becomes acceptable (over 60%).
c) Results from each laboratory separately, each replicated set combined (whether or not they pass the χ^2 test), and the two laboratories together, all results of the same replication (within- and between-laboratories) combined.
d) As above, but outliers (i.e. individual measurements that cause the replicated set to fail the χ^2 test) removed. Although one replicated set was consistently flagged as having a poor fit with the rest of the sequence (T18167 = RT3949; a seeds sample from Megiddo H5 [=VB − IVA]) was retained since the overall agreement of the model is acceptable.

All the models gave broadly similar results, with the peaks of the boundary distribution varying in the order of magnitude of 20–25 yr between the different model runs, and the limits of the 68% and 95% highest-density regions slightly more (Table 4).

Table 4 Iron I|IIA transition calculated in different ways.[a]

	Tucson	Rehovot	Rehovot and Tucson
Not combined with misfits	900BCE 926—881 950————830 3 misfits: T18154a, T18154aa, T18167aa	875BCE 898—859 960————850 6 misfits: 3932.5–6, 3786.4, 3783.4, 3949.3–4	890BCE 912 ⊔—866 940————850 7 misfits: 3932.5–6, 3786.4, 3783.4, T18167aa, 3949.3–4
Not combined after the removal of misfits until the overall agreement is >60%	900BCE 926—881 950————830 3 misfits: T18154a, T18154aa, T18167aa	890BCE 908—866 925————835 3 misfits left: 3932.5, 3783.5, 3949.3	890BCE 913—882 930————845 5 misfits left: 3932.5, 3783.4–5, T18167aa, 3949.3
Combined with outliers	890BCE 950—880 980————840 No outliers 1 misfit: T18167	880BCE 891—861 920————840 2 outliers: 3803.5, 3786.4 1 misfit: 3949	910BCE 930—860 960————840 4 outliers: T18161aa, 3803.5, 3786.4, 3783.4 1 misfit: T18167 = 3949
Combined after the removal of outliers	890BCE 950—880 980————840 1 misfit: T18167	880BCE 892—852 960⊔ —————840 1 misfit: 3949	870BCE 900—840 960————830 1 misfit: T18167 = 3949

[a]The dates in italicized numerals at the top of each plot are the maximum likelihood estimate; the dates in the smaller range immediately below the plot are the 68% highest-density interval, and dates in non-italicized numerals in the bottom range are the 95% highest-density interval.

CONCLUSIONS

The ^{14}C intercomparison exercise presented here is one of the most comprehensive to have been run between laboratories on a substantial number of real archaeological samples. Its implications must be assessed at different levels.

Bias or No Bias?

The reproducibility between the Tucson and Rehovot laboratories as presented here is as close as one could expect, and certainly well within the standards of the FIRI exercise. No bias can be detected either between the laboratories, between different pretreatment procedures, or between measuring techniques. Given enough measurements (see below), the outcome regarding the object of inquiry here, the Iron I|II transition, would have, *grosso modo*, been the same had we confined ourselves to one lab/method only. This is a prerequisite in the framework of our more comprehensive research, which, as mentioned, involves 400 or so measurements. For the time being, many measurements are made by one method/laboratory only, though we are currently expanding the intercomparison program to include not only more samples and other laboratories, but also more detailed characterization of the materials used for ^{14}C analysis. A detailed study of different preparation procedures used in different laboratories may shed light on the reasons for the still existing outliers.

Can [14]C Dating Resolve the Levantine Iron Age Dilemma?

The obvious answer here is yes, provided that enough measurements are available. Given the current state-of-the-art technology in archaeological seriation, ^{14}C accuracy and precision, and statistical modeling, the investigation of the chronology of the eastern Mediterranean in this period will have to be based on numerous, replicated dates, taken from different sites and measured by different procedures. A large, replicated data set is the only way to overcome the inevitable noise in the model.

The Iron I|II transition dates as calculated here range between 910–875 BCE, diverging only by 10–25 yr between the 2 labs. They are about 30 yr higher than the transition dates calculated based on the Dor dates alone (880–850, e.g. Gilboa and Sharon 2003: Table 21)—another issue awaiting investigation. However, they provide further empirical support for a lower chronology.

We must stress again that we do not consider the transition dates presented here as the final conclusion of our research for 3 main reasons: a) they were calculated using only the measurements participating in the intercomparison exercise—a small fraction of the forthcoming ~400 dates we have measured; b) clustering the dates into (only) 2 very broad archaeological horizons (Iron I and Iron II), as was done here, is not sufficient for the typological resolution this problem requires; c) the data set presented in this paper lacks dates from the horizon that we termed "Iron I|II." This, and a much more detailed, relative placement of all the dates in the Iron I–II continuum will be considered when all of our results are taken into account.

ACKNOWLEDGMENTS

REFERENCES

Bevington PR, Robinson DK. 1992. *Data Reduction and Error Analysis for the Physical Sciences*. Boston: WCB/McGraw-Hill.

Bland JM, Altman DG. 1986. Statistical method for assessing agreement between two methods of clinical measurement. *The Lancet* 1:307–10.

Bronk Ramsey C. 1994. Analysis of chronological information and radiocarbon calibration, the program OxCal. *Archaeological Computing Newsletter* 41:11–6.

Bronk Ramsey C. 1995. Radiocarbon calibration and the analysis of stratigraphy, the OxCal program. *Radiocarbon* 37(2):425–30.

Bronk Ramsey C. 2000. Comment on "The use of Bayesian statistics for ^{14}C dates of chronologically ordered samples: a critical analysis." *Radiocarbon* 42(2):199–202.

Bruins H, van der Plicht J, Mazar A. 2003. Response to comment on "[14]C dates from Tel Rehov: Iron-Age chronology, pharaohs and Hebrew kings." *Science* 302:568c.

Coldstream JN. 2003. Some Aegean reactions to the chronological debate in the southern Levant. *Tel Aviv* 30:247–58.

Coldstream JN, Mazar A. 2003. Greek pottery from Tel Rehov and Iron Age chronology. *Israel Exploration Journal* 53:29–48.

Fantalkin A. 2001. Low Chronology and the Greek protogeometric and geometric pottery in the southern Levant. *Levant* 33:117–25.

Finkelstein I. 1996. The archaeology of the United Monarchy: an alternative view. *Levant* 27:177–87.

Finkelstein I, Piasetzky E. 2003. Wrong and right; high and low: [14]C dates from Tel Rehov and Iron Age chronology. *Tel Aviv* 30:283–95.

Gilboa A, Sharon I. 2001. Early Iron Age radiometric dates from Tel Dor: preliminary implications for Phoenicia, and beyond. *Radiocarbon* 43(3):1343–51.

Gilboa A, Sharon I. 2003. An archaeological contribution to the early Iron Age chronological debate: alternative chronologies for Phoenicia and their effects on the Levant, Cyprus and Greece. *Bulletin of the American Schools of Oriental Research* 332:7–80.

Holden C. 2003. Dates boost conventional wisdom about Solomon's splendor. *Science* 300:229–31.

Kopcke G. 2002. 1000 BCE? 900 BCE? A Greek vase from Lake Galilee. In: Ehrenberg E, editor. *Leaving No Stones Unturned: Essays on the Ancient Near East and Egypt in Honor of Donald P Hansen*. Winona Lake: Eisenbrauns. p 109–17.

Mazar A. 2004. Greek and Levantine Iron Age chronology: a rejoinder. *Israel Exploration Journal* 54:24–36.

Scott EM. 2003. The Fourth International Radiocarbon Intercomparison (FIRI). *Radiocarbon* 45(2):35–150.

Sharon I. 2001. "Transition dating"—a heuristic mathematical approach to the collation of ^{14}C dates from stratified sequences. *Radiocarbon* 43(3):345–54.

Sharon I, Gilboa A, Boaretto E. Forthcoming. [14]C and the early Iron Age of Israel—Where are we really at? A commentary on the Tel Rehov radiometric dates. *Proceedings of the 2nd EuroConference SCIEM 2000*. Vienna: Austrian Academy of Sciences.