THE UNIVERSITY OF
WARWICK

Original citation:
Smith, Oliver, Clapham, Alan J., Rose, Pam, Liu, Yuan, Wang, Jun and Allaby, Robin G.. (2014) A complete ancient RNA genome : identification, reconstruction and evolutionary history of archaeological Barley Stripe Mosaic Virus. Scientific Reports, Volume 4 . Article number 4003. ISSN 2045-2322

Permanent WRAP url:
http://wrap.warwick.ac.uk/59386

Copyright and reuse:
The Warwick Research Archive Portal (WRAP) makes this work of researchers of the University of Warwick available open access under the following conditions.

This article is made available under the Creative Commons Attribution 3.0 (CC BY 3.0) license and may be reused according to the conditions of the license. For more details see: http://creativecommons.org/licenses/by/3.0/

A note on versions:
The version presented in WRAP is the published version, or, version of record, and may be cited as it appears here.

For more information, please contact the WRAP Team at: publications@warwick.ac.uk
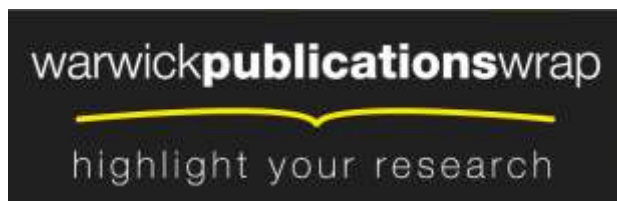
warwick**publications**wrap
highlight your research

http://wrap.warwick.ac.uk

Correspondence and
requests for materials
should be addressed to
R.G.A. (r.g.allaby@
warwick.ac.uk)

A complete ancient RNA genome: identification, reconstruction and evolutionary history of archaeological Barley Stripe Mosaic Virus

Oliver Smith[1], Alan Clapham[1], Pam Rose[2], Yuan Liu[3], Jun Wang[4] & Robin G. Allaby[1]

[1]School of Life Sciences, Gibbet Hill Campus, University of Warwick, Coventry CV4 7AL, [2]The Austrian Archaeological Institute; Cairo Branch, Zamalek, Sharia Ismail Muhammed, Apt 62/72, Cairo, Egypt, [3]BGI-Europe-UK, 9 Devonshire Square, London, EC2M 4YF, UK, [4]BGI-Shenzhen, Shenzhen 518083, China.

The origins of many plant diseases appear to be recent and associated with the rise of domestication, the spread of agriculture or recent global movements of crops. Distinguishing between these possibilities is problematic because of the difficulty of determining rates of molecular evolution over short time frames. Heterochronous approaches using recent and historical samples show that plant viruses exhibit highly variable and often rapid rates of molecular evolution. The accuracy of estimated evolution rates and age of origin can be greatly improved with the inclusion of older molecular data from archaeological material. Here we present the first reconstruction of an archaeological RNA genome, which is of Barley Stripe Mosaic Virus (BSMV) isolated from barley grain ~750 years of age. Phylogenetic analysis of BSMV that includes this genome indicates the divergence of BSMV and its closest relative prior to this time, most likely around 2000 years ago. However, exclusion of the archaeological data results in an apparently much more recent origin of the virus that postdates even the archaeological sample. We conclude that this viral lineage originated in the Near East or North Africa, and spread to North America and East Asia with their hosts along historical trade routes.

T he evolution of domestication is associated with the rise of many diseases attributed to an agricultural origin[1]. The unprecedented population densities of humans, domesticated animals and plants in which efficient transmission rates were possible provided new pools for disease. This combined with the novel juxtaposition of species as domesticates came into contact with humans, each other and indigenous wild species in new environments facilitated the transfer of diseases between species, often with an associated increased virulence in the adopted host. While this process has long been appreciated as an origin of many human diseases[1], more recently it has become apparent that the origins of many domesticated plant diseases are recent, and can be categorized into three principal time periods of origin[2–4]. Firstly, many diseases may be associated with the origins of agriculture as increased plant stand densities and intensifying agricultural practices caused both the amplification of existing diseases from the wild progenitors as well as transmission from other wild species of the centre of crop origin for crops such as wheat, maize and rice[2–9]. Secondly, subsequent to domestication, the spread to new environments with agricultural expansion caused domesticated plants to come into contact with new indigenous wild populations resulting in host transfer to crops in the past few thousand years such as the transfer of pathogens from wild European orchids to Brassicas 1,000 years ago[10,11]. Thirdly, more recent global movements of plants and disease vectors in the past few hundred years have also caused the emergence of significant pathogens from wild hosts from quite disparate geographies, such as the Israeli origin of Tomato yellow leaf virus which adopted tomato, an Andean crop, as a host that subsequently spread to the Americas, Japan and Australia[2].

Emerging infectious diseases of plants are recognized as a growing threat to global food security, and of those viruses account for almost half[12]. Therefore a better understanding of the origins of diseases is of significant importance to the management of global food resources. Ancient DNA and RNA of viruses obtained directly from herbaria and long-term field sampling have demonstrated that heterochronous sampling serves to improve phylogenetic based estimates by countering recent calibration bias and often resulting in a greater time depth for

the estimate of viral origins[4,5,13–18]. However, the oldest specimens that have been used to date have been around 100–150 years in age[13,17]. It is therefore possible that further improvements on the estimate of age of virus origins could be obtained from viruses recovered from older, archaeological material.

The most studied plant virus that infects multiple crops is Barley Stripe Mosaic Virus (BSMV)[19]. BSMV is a tripartite, positive-strand RNA virus of the *Hordeivirus* genus, which has no vector and is transmitted through direct contact via pollen and seeds[20]. It is important both in terms of the economic damage it incurs in wheat and barley crops[21], but also because it has become a tool for studying gene function in plants *in vivo* by virus-induced gene silencing methods. Furthermore, BSMV is capable of infecting hundreds of other members of the Poaceae including *Zea mays*, as well as some eudicots such as *Nicotinia* and *Chenopodium*[22]. The virus was discovered in 1950, and the earliest record of the symptoms associated with it was only 100 years ago[23]. Subsequently, BSMV has been found in North America, Europe and Asia and more recently in North Africa[24,25]. This disease therefore has the appearance of a recent origin and rapid spread. We were able to test this hypothesis by screening archaeological samples of barley spanning a range of ages from a few hundred to three thousand years for the presence of the virus, and used next generation sequencing to retrieve archaeogenomic levels of ancient RNA in order to attempt a genomic reconstruction.

Results and Discussion

Presence of BSMV in archaeological barley. The detection of ancient BSMV virus requires the preservation of ancient RNA in the archaeological record. To date there have been few studies on RNA[26,27] because it is highly labile relative to DNA, with an expected 50-fold increased rate of degradation[28]. RNA is highly vulnerable to hydrolytic attack relative to DNA, but less prone to depurination[29,30]. Therefore in very dry environments, the rate of RNA degradation may be greatly reduced[30]. Archaeological barley samples were selected from Qasr Ibrim, a site in North Africa where BSMV has been reported. The arid conditions of the site are expected to be conducive to RNA preservation as has been observed for other biomolecules in samples of comparable age[31–33]. We tested grains of barley from Islamic, Late Christian, Meroitic and Napatan archaeological strata by RT-PCR of one target from each of the three genomic segments: the helicase gene (alpha segment; 123 nt), the beta-C gene (beta segment; 126 nt) and the RNA-dependent RNA polymerase gene (gamma segment; 137 nt), table S1. Amplicons were successfully produced only from the Late Christian sample (QI84/71) that is archaeologically dated to 600–900 years BP. This sample was subject to an archaeogenomic analysis of its small RNA fraction.

Genome reconstruction and authentication of an archaeological BSMV virus. The small RNA fraction of sample QI84/71 was sequenced on the Illumina Hi-Seq platform. Allowing 1 mismatch per read of average length 20 nt, we assigned 49,232 non-redundant reads and 133,756 individual reads as BSMV. These were aligned to a reference genome in 51 contigs over 10,171 nucleotides with an average coverage depth of 300 reads and cross-genome coverage of 99.4%. From this data we constructed a consensus archaeological genome of the BSMV virus for further analysis. As a control we checked that the genome we had generated did not in any way reflect the Barley genome content. This may be expected from a retrovirus but was not expected from a *Hordeivirus*. *In silico* fragmentation (see methods) and Bowtie alignment of the archaeological virus genome to the barley (Morex) genome[34] showed no continuous sequences of >50 nt present in the barley genome, even when allowing the maximum 3 mismatches under alignment parameters. We therefore conclude that the

archaeological genome does not represent an unexpected form of barley genome incorporation of the virus.

While historical virus RNA of less than 100 years in age has been authenticated[35–37], the conclusions of some previous studies of potentially older aRNA virus data have been attributed to contamination[38,39]. For further authentication we employed a method of comparative base modification profiles to establish the likelihood of this being a modern contaminant strain. In aDNA, one of the most common postmortem degradation markers is that of cytosine to uracil deamination by hydrolysis of cytosine by H_2O[40–42]. We expected to observe a similar process in RNA as has been suggested in studies of RNA editing[43]. The proportion of $C > U$ differences relative to other base change types of individual reads from the archaeological sample mapped to the BSMV genome contrasted significantly to that observed between published modern BSMV virus sequences and the consensus BSMV genome sequence, figure S1. We then compared the linear regression values of base changes between modern and archaeological strains, both with ($R^2 = 0.647$) and without ($R^2 = 0.934$) $C > U$ transitions. We also performed an equality of variance F-test for both complete modification profiles (p = 0.12), showing that $C > U$ transitions are more prevalent in the Qasr Ibrim strain than would be expected from *in vivo* processes alone. We confirmed likelihood of these transitions being the result of postmortem degradation by running BAM alignments of each segment against a reference strain using mapDamage v2.0[44]. Interestingly, the distribution of probable postmortem $C > U$ transitions was more pronounced at 5′ and 3′ ends (figure S1c) where comparable DNA studies involve removal of 3′-overhangs as part of library preparation[44]. This could be due to the central portion of the single stranded RNA molecule being protected from hydrolytic attack through secondary structure leaving both ends overhanging and so more prone to deamination.

This evidence leads us to conclude that we have detected a genuine instance of BSMV from a ~750-year-old archaeological sample of barley, which is the oldest authenticated instance of an ancient virus, and the first archaeological virus genome that has been sequenced.

The age of origin of BSMV. In order to reconstruct the evolutionary history of BSMV, we compared the archaeological genome sequence with six modern, complete genome sequences that have been published over the past 3 decades[45–50]. A previous study suggested a phylogenetic mosaicism between BSMV and the other two members of the *Hordeivirus* genus, poa semilatent virus (PSLV) and lychnis ringspot virus (LRSV), at different gene loci[20]. It would make the use of the whole genome in a phylogenetic analysis problematic if the different genes represented very different phylogenetic histories associated with separate species genomes. However, we found that PSLV consistently formed the closest sister taxon to BSMV at all seven genes that occurred in all three viruses, figure S2. We conclude that the evidence does not support a history of recombination between BSMV and LRSV. Therefore we selected PSLV as the outgroup in all our subsequent phylogenetic analyses.

We used BEAST to construct a heterochronous calibrated tree both in the presence and the absence of the archaeological BSMV sample, Figure 1 and Figure S3. A constant population size model was applied when comparing BSMV and PSLV as has been recommended for interspecific comparisons[51]. We applied an epidemiological model when analyzing diversity between different BSMV strains as has been recommended for intraspecific viral comparisons[52], although we found little difference between these results and the application of a constant population model. We compared the marginal likelihood of both models by computing the Bayes factor through posterior distribution sampling. Since the harmonic mean of the sampled likelihoods is an estimator of the marginal likelihood, we calculated the harmonic mean of the likelihood from the posterior output of each model and extrapolated the difference in
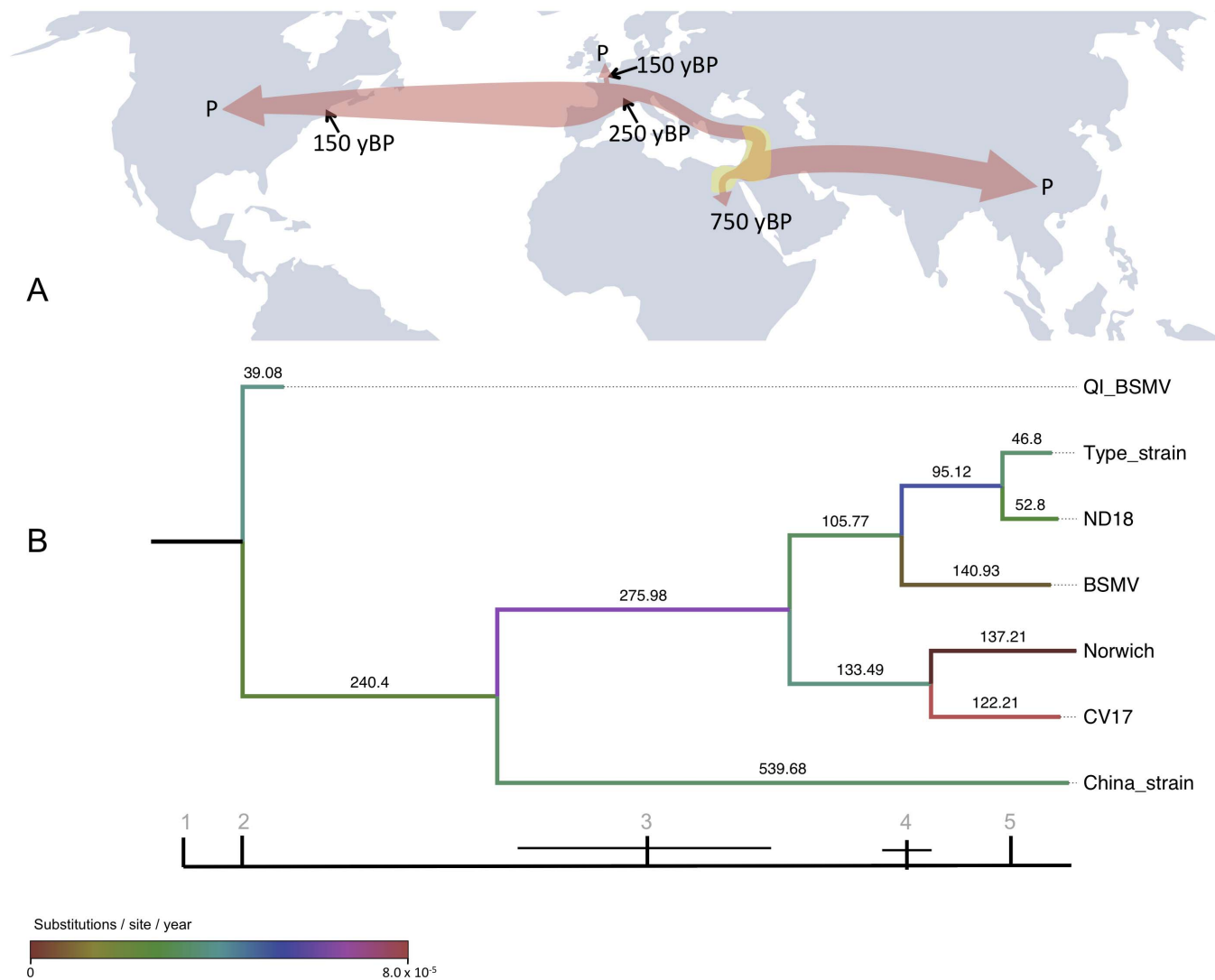
Figure 1 | Hypothesised population movements and phylogeny of BSMV. (A): Global distribution map showing likely Fertile Crescent origin of BSMV ~100,000 yBP - present. The single yellow circle indicates probable area of origin of BSMV distribution prior to infection of Qasr Ibrim. Red arrows indicate probable transmission routes based on available strain data; West to East across Europe and to North America during the Industrial and Green Revolutions, and East to West along the Silk Road to China and East Asia during the medieval period. 'P' indicated present day and applies globally as BSMV is thought now to be found worldwide. Map template provided by (http://www.freeworldmaps.net/pdf/maps.html). (B): BEAST analysis of all extant and Qasr Ibrim BSMV strains using an epidemiology model. The topology of BSMV using this model is consistent with that including an outlier (see Figure S3) with similar node dating and consistent clustering of western strains compared to archaeological or eastern strains. Relevant events indicated on timeline as follows. 1: MRCA of BSMV strains ~10,000 yBP (timeline unscaled). 2: Projected initial infection of BSMV at Qasr Ibrim, possibly terminated during the Late Christian era. 3: Re-opening of Silk Road trade route by Mongol Empire allowing the spread of BSMV into Asia. 4: Industrial revolution in Europe and the USA being a potential gateway from Europe to North America. 5: Postwar global Green Revolution; intensive farming providing greater opportunity for viral transmission and spread. Note variability in molecular evolution rates (branch colours denoted by key) with increased rates occurring at the root of western strains and more recently along the CV17 and Norwich strains.

log space. We found the difference is too small to be significant so we did not reject the epidemiological model which also had the greater mean likelihood of the two models albeit marginally, Figure S5c.

We found that the inclusion of the archaeological BSMV virus calibration had a striking impact on the estimate of the age of the origin of the virus, a phenomenon reported previously given the inclusion of actual or virtual recalibration[37,53]. In the absence of the archaeological virus calibration, we estimate a mean substitution rate of 7.3×10^{-4} substitutions/site/year, figure S4. The base of the crown group dates to about 50 years ago, and the BSMV appears to split from PSLV about 150 years ago, figure S3. This data would seem to support the recent origin of the virus in keeping with its recent discovery and documented evidence of BSMV-like symptoms. However, our discovery of the virus in 750-year-old archaeological

samples clearly challenges that scenario. The phylogenetic reconstruction using the archaeological virus places it close to the base of the crown group, suggesting that this sample was close to the origin of the spread of modern BSMV. The plesiomorphic position is supported by the lower number of substitutions on the branch leading to the archaeological virus, which further validates its authenticity, figure 1. In this reconstruction we find a range of substitution rates that are considerably lower than if the archaeological data are excluded (Figure S4). We find the common ancestor of BSMV and PSLV occurs within a large temporal range, the median date of the range correlates to the time of domesticated barley's appearance in the archaeological record[54], Figure S5b. To explore the information value underlying the wide temporal range we plotted the frequency distribution of branch ages leading to PSLV from the

origin, Figure S5d. We observe a strong modal signal at 2000 years BP, with 80% of the branches being less than 28000 years old. The age of the common ancestor is therefore likely to be in the younger part of the temporal range estimated from BEAST. It is possible that further instances of BSMV may be present earlier in the archaeological record and would alter the tree calibration if found as has been reported previously[37].

Evolution and spread of BSMV. The archaeological BSMV genome is close to the base of the crown of the phylogeny indicating that it is closely associated with the origin of modern populations of the virus. Interestingly, this dates to around to 1234 AD in the phylogenetic reconstruction, which corresponds closely to the seventh Crusade of Louix IX[55]. While such dating should be interpreted with caution because of the naturally associated wide margins of error (figure S6), it is interesting to note that this was one of the few Crusade episodes to visit North Africa, giving some credibility that this episode could have been instrumental in spreading the virus to Europe. The tree topology shows a clear divide between a single Chinese lineage, and the remaining lineages that are associated with Europe and America. This split in east and west lineages dates to around the 1473 AD, which is a century after the Mongol Empire stabilized the Silk Road. It is likely that BSMV was transported to the east via trade routes such as the Silk Road in the late Medieval period. While we can estimate that BSMV left for East Asia just over 500 years ago, in the absence of further East Asian BSMV genome sequences we cannot estimate the time that the virus arrived.

We observed variability in the substitution rate associated with the expansion of BSMV (Figure 1) a finding similar to that of a previous palaeoepidemiological study of the plague[56]. Notably, the initial branch leading to the European and American lineages shows an increased rate of 7.01×10^{-5} substitutions/site/year that may be associated with an early epidemic of the virus from the 15th–18th centuries. A notable further rate increase was observed in the CV17 lineage in recent years. The Chinese lineage on the other hand shows no increase in rate that may indicate no particular epidemiological expansion in BSMV, perhaps reflecting the lower emphasis on barley in agriculture in the east. In more recent history, the virus appears to have spread to the US from Europe around 120–150 years ago whereupon an increased rate of substitution occurred, possibly caused by either a second epidemiological expansion or increased environmental or human-mediated selection pressure.

To investigate how the virus had changed functionally since the Late Christian period in North Africa we examined the non-synonymous substitutions that differentiated the archaeological BSMV from modern virus genomes. We identified 14 amino acid substitutions (table S2), none of which were obviously attributable to a change in function. Interestingly, most changes were located in the methyltransferase and polymerase coding genes, while only three occurred in other genes. Tertiary structure prediction of coding gene products using the SWISS-MODEL workspace showed no significant change in any identifiable domain when compared to the gene product of the BSMV reference strain[45], where structure prediction was possible. Conserved amino acid positions in coding genes were identified by comparison with the wider gene family using pFam[57]. The archaeological sequence showed no differences in conserved positions to the BMSV reference strain. The archaeological genome therefore appears to represent a functional genome sequence, but it is unknown whether this virus would have had any difference in virulence to modern BSMV.

To survey for evidence of selection in the BSMV genome over the course of its history we examined the ratio of non-synonymous to synonymous SNPs (dN/dS) for each branch of the phylogenetic tree, Figure S6 and S7, and Table S2. The overall dN/dS ratio of the tree is high relative to the range expected for plant viral evolution[57], which could be indicative of adaptive evolution across the BSMV tree.

However, the proportions of nonsynonymous SNPs for Qasr Ibrim are low both compared to the other BMSV strains and the overall average expected for viruses. We conclude that it is unlikely that significant adaptive evolution occurred along this lineage. In contrast, the remaining strains show evidence of several episodes of adaptive evolution. The most significant dN/dS signal of adaptive change was found on the CV17 lineage, which was also associated with the greatest increase in the base substitution rate we observed in Figure 1. This is the only lineage to be highly significant relative to the other BSMV strains. However, all but two branches indicate a significant signal of selection when compared to expected plant viral range for dN/dS ratios. Generally, the strength of signal does not correlate closely with episodes of increased base substitution rates, rather a picture emerges of continual adaptive evolution over the history of the virus.

An archaeogenomic perspective of BSMV origin. The calibration introduced by the inclusion of an archaeological virus profoundly alters the interpretation of the likely origin of BSMV. Our estimation of the BSMV rate of molecular evolution based on the archaeological data is approximately 3.9×10^{-5} substitutions per site per year, a slower rate than estimated by recent serial and heterochronous sampling. This study demonstrates that a virus that has the appearance of a recent origin, which could lead to policy decisions about movement of plant material, is shown to have a much older origin associated with historic or prehistoric activity. It is likely that early barley first contracted BSMV from a local wild grass population. Hordeiviruses rarely jump hosts because of a lack of vectors involved in their transmission. Consequently, direct contact between grains of the crop and grass weeds caught in cultivation is a likely mechanism of transmission and consequently origin in this case. It is possible that other viruses that similarly appear to be of very recent origin may in fact have a more ancient origin.

The inclusion of archaeological data allows the observation of genomic change over a relatively long time scale which includes periods of variable and adaptive evolution rates based on population movement, infection of new cultivars and changing environmental conditions according to those new hosts. These data point to a phylogeographic expansion of BSMV beginning around the time of the European Industrial Revolution and continuing into the 21st century. Our data appear to indicate that until around 800 years ago, BMSV did not become widespread. We cannot estimate with any precision how long BMSV existed in the agricultural system before this expansion from these data. It may have crossed into barley during the initial domestication process, although a later entry a couple of thousand years ago into barley is more strongly supported by the data. It may be the case that BMSV did not become well established in agriculture until a threshold of intensity had been crossed in terms of density of barley stands and field sizes. In this case it could have been the demands on the local agricultural economy of the medieval war machines that required an extensive associated supporting population aside from the armies themselves that caused such a threshold to be crossed.

Methods

manufacturer's instructions. Final RNA extractions were quantified using the Qubit broad-range RNA assay (Invitrogen).

Illumina library construction. Illumina libraries were built using a Small RNA Library Preparation Kit (Bioo Scientific) according to manufacturer's instructions. All pre-PCR steps of library construction from archaeological RNA took place in the dedicated clean facility. Sequencing was performed using the HiSeq 2500 platform (Illumina) using a single lane for the sample.

Metagenomics. FastQ data was converted to FastA format and redundant sequences consolidated using scripts of the author's design, retaining frequency information within FastA tags. All sequences of abundance >1 were subject to BLASTn analysis using a custom 64-bit build of version 2.2.25 on a dedicated server. BLASTn output files were parsed for metagenomic data using MeGAn, where the presence of BSMV RNA was first identified in the archaeological dataset.

Bowtie alignment. Detection of all potential BSMV reads from archaeological data was performed using Bowtie version 0.21. Bowtie indexes were constructed from the original BSMV reference strain (Genbank BioProject PRJNA15031) for each of three chromosomal segments. Alignment was performed using complete Illumina sequence data in consolidated FastA format, allowing for one mismatch per read and so constraining alignments to >95% accuracy.

Control check against Morex genome. To check that the viral genome that was constructed was not an artifact from the host barley genome, the first was checked against the latter for regions of similarity. The viral genome sequence was randomly fragmented 3 times into 100, 50 and 18–21 nt sections (*in silico* fragmentation). An alignment was then attempted of each set of fragments against the Morex genome using Bowtie. No fragments of 50 nt or larger aligned, 129 fragments of 18–21 nt section aligned to the Morex genome accounting for 40 contiguous sequences of < 28 nt length and representing 9.7% of the total BSMV genome.

Authentication by nucleotide deamination profiles. Archaeological reads positively aligned to the BSMV reference strain by Bowtie were parsed according to alignment specificity. To avoid statistical bias arising from high-frequency reads containing *in vivo* polymorphisms, base modification types of all non-redundant reads were collated from Bowtie output files and summarised, using a Unix shell script of the author's design. Further authentication of deamination was carried out with mapDamage 2.0 using standard parameters[44].

Genome assembly. Reads identified using these methods were aligned to the full BSMV genome in segments to visualize coverage depth using Geneious v5.1. 96.6% coverage was achieved across 51 intervals. The draft archaeological genome was aligned to 6 available BSMV reference strains, each comprising 3 segments (Genbank accessions NC_003469.1, NC_003481.1, NC_003478.1, J04342.1, X03854.1, M16576.1, AY789693.1, AY789694.1, AY787207.1, U35767.1, U35770.1, U13917.1, U35766.1, U35769.1, U13916.1, JF803283.1, JF803284.1, JF803285.1) using the Geneious align feature. For a second draft, sequence gaps were putatively filled first by re-alignment of short reads to the BSMV genome, allowing 3 mismatches as opposed to 1, giving 99.4% coverage. Where gaps existed after this alignment, the 62 remaining character states were inferred based on maximum likelihood character state reconstruction using Mesquite from the topology resulting from BEAST analysis of the first draft genome.

Preliminary phylogenetic analysis. For nucleotide-based phylogenetic analysis, segments of the tripartite genome were concatenated into a single contiguous genome. Sequence gaps within the archaeological genome and corresponding bases in aligned genomes were removed so as to avoid bias in subsequent phylogenetic assignment. The closest relative within the *Hordeivirus* genus, Poa semilatent virus (PSLV) was chosen as an outlier and also subject to gap/nonaligned segment removal. Neighbor-joining trees were constructed based on this initial alignment to ascertain a guide phylogeny to preclude incorrect clade assignment as a result of using tip dates with BEAST.

Phylogenetic analysis using Bayesian inference method. Phylogenetic placement based on available sequence data for 6 strains of BSMV and corresponding sections of PSLV as an outgroup was performed using BEAST using a relaxed lognormal clock and epidemiology tree prior. All BSMV strains were partitioned into a single monophyletic group for comparison to the outlier. Tip dates for each accession were defined by date of publication for extant strains and estimated as mid-period (Late Christian; 600–900 years BP) for the archaeological being 750 years old. A comparison Neighbor-joining tree was constructed with Geneious using the same alignment.

To preclude the possibility of given tip dates resulting in bias of branch lengths, manual calculation of relative branch lengths based on SNPs of the BSMV/PSLV alignment (3,351 bases) was performed by observing SNPs according to a branch model output by tip-dated BEAST analysis. The number of SNPs along each branch was used to calculate relative branch lengths.

Calculation of molecular evolution rate. Molecular evolution rates of BEAST outputs from BSMV alignments only, and separate analysis including the PSLV outlier, were calculated using Tracer. Sequence gaps in the archaeological sample,

their corresponding bases in extant strains and ambiguous bases in all strains were discarded so as not to bias analysis.

Evaluation of potential selection pressures. Selective pressures identifiable though nonsynonymous SNPs (nsSNPs) were assessed across full genomes of 7 strains (6 full extant strains and the archaeological strain). Expected frequencies of nsSNPs for each strain (E) were calculated according to the probability of obtaining the observed number (O) of snSNPs based on the total length and the global frequency of nonsynonymous SNPs. The Chi squared (χ^2) statistic was used to calculate significance of deviation of observed and expected values.

1. Wolfe, N. D., Dunavan, C. P. & Diamond, J. Origins of major human infectious diseases. *Nature* **447**, 279–283 (2007).
2. Jones, R. A. C. Plant virus emergence and evolution: Origins, new encounter scenarios, factors driving emergence, effects of changing world conditions, and prospects for control. *Virus Res.* **141**, 113–130 (2009).
3. Stukenbrock, E. H. & McDonald, B. A. The origins of plant pathogens in agro-ecosystems. *Ann. Rev. Phytopath.* **46**, 75–100 (2008).
4. Gibbs, A. J., Fargette, D., Garcia-Arenal, F. & Gibbs, M. J. Time - the emerging dimension of plant virus studies. *J. Gen. Virol.* **91**, 13–22 (2010).
5. Fargette, D. *et al.* Diversification of rice yellow mottle virus and related viruses spans the history of agriculture from the Neolithic to the present. *PLoS Pathog.* **4**, e1000125 (2008).
6. Gibbs, A. J., Ohshima, K., Phillips, M. J. & Gibbs, M. J. The Prehistory of Potyviruses: their initial radiation was during the dawn of agriculture. *Plos One* **3**, e2523 (2008).
7. Stukenbrock, E. H., Banke, S. R., Javan-Nikkhah, M. & McDonald, B. A. Origin and domestication of the fungal wheat pathogen Mycosphaerella graminicola via sympatric speciation. *Mol. Biol. Evol.* **24**, 398–411 (2007).
8. Couch, B. C. *et al.* Origins of host-specific populations of the blast pathogen Magnaporthe oryzae in crop domestication with subsequent expansion of pandemic clones on rice and weeds of rice. *Genetics* **170**, 613–630 (2005).
9. Munkacsi, A. B., Stoxen, S. & May, G. Domestication of maize, sorghum, and sugarcane did not drive the divergence of their smut pathogens. *Evolution* **61**, 388–403 (2007).
10. Nguyen, H. D. *et al.* Turnip mosaic potyvirus probably first spread to Eurasian Brassica crops from wild orchids about 1000 years ago. *Plos One* **8**, e55336; DOI: 10.1371/journal.pone.0055336 (2013).
11. Brunner, P. C., Schurch, S. & McDonald, B. A. The origin and colonization history of the barley scald pathogen *Rhynchosporium secalis*. *J. Evol. Biol.* **20**, 1311–1321 (2007).
12. Anderson, P. *et al.* Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends Ecol. Evol.* **19**, 535–544 (2004).
13. Fraile, A. *et al.* A century of tobamovirus evolution in an Australian population of Nicotiana glauca. *J. Virol.* **71**, 8316–8320 (1997).
14. Almeida, R. P. P., Bennett, G. M., Anhalt, M. D., Tsai, C. W. & O'Grady, P. Spread of an introduced vector-borne banana virus in Hawaii. *Mol. Ecol.* **18**, 136–146 (2009).
15. Duffy, S. & Holmes, E. C. Phylogenetic evidence for rapid rates of molecular evolution in the single-stranded DNA begomovirus tomato yellow leaf curl virus. *J. Virol.* **82**, 957–965 (2008).
16. Duffy, S. & Holmes, E. C. Validation of high rates of nucleotide substitution in geminiviruses: phylogenetic evidence from East African cassava mosaic viruses. *J. Gen. Virol.* **90**, 1539–1547 (2009).
17. Malmstrom, C. M., Shu, R., Linton, E. W., Newton, L. A. & Cook, M. A. Barley yellow dwarf viruses (BYDVs) preserved in herbarium specimens illuminate historical disease ecology of invasive and native grasses. *J. Ecol.* **95**, 1153–1166 (2007).
18. Simmons, H. E., Holmes, E. C. & Stephenson, A. G. Rapid evolutionary dynamics of zucchini yellow mosaic virus. *J. Gen. Virol.* **89**, 1081–1085 (2008).
19. Lee, W. S., Hammond-Kosack, K. E. & Kanyuka, K. *Barley Stripe Mosaic Virus* – mediated tools for investigating gene function in cereal plants and their pathogens: virus-induced gene silencing, host-mediated gene silencing, and virus-mediated overexpression of heterologous protein. *Plant Physiol.* **160**, 582–590 (2012).
20. Jackson, A. O., Lim, H. S., Bragg, J. Ganesan. & Lee, M. Y. Hordeivirus replication, movement and pathogenesis. *Ann. Rev. Phytopath.* **47**, 385–422 (2009).
21. Carroll, T. Barley stripe mosaic virus: its economic importance and control in Montana. *Plant Disease* **64**, 136–140 (1980).
22. Jackson, A. O. & Lane, L. C. Hordeiviruses. In Kurstak, E. ed, *Handbook of Plant Virus Infections and Comparative Diagnosis.* Elsevier, Amsterdam, pp 565–625 (1981).
23. McKinney, H. H. Biological characteristics of barley stripe mosaic virus strains and their evolution. *USDA Agricultural Research Service Technical Bulletin.* **1324**, 1–79 (1965).
24. Niffenegger, D., Mills, I. K. & Davis, D. J. Effects of Barley Stripe Mosaic Virus upon growth of infected barley plants. *Crop Science* **5**, 227–229 (1965).
25. Hafez, E. E., Abdel Aleem, E. E. & Fattouh, F. A. Comparison of barley stripe mosaic virus strains. *Zeitschrift fur Naturforschung. C, J. Biosci.* **63**, 271–276 (2008).

26. Rollo, F. Characterisation by molecular hybridization of RNA fragments isolated from ancient (1400 B.C.) seeds. *Theor. Appl. Genet.* **71**, 330–333 (1985).

27. Fordyce, S. L. *et al.* Deep sequencing of RNA from ancient maize kernels. *Plos One* **8**, e50961 (2013).

28. Eigner, J., Boedtker, H. & Michaels, G. The thermal degradation of nucleic acids. *Biochim. Biophys. Acta* **51**, 165–168 (1961).

29. Thompson, J. E. *et al.* Limits to catalysis by ribonuclease A. *Bioorg. Chem.* **23**, 471–481 (1995).

30. Fordyce, S. L. *et al.* Long-term RNA persistence in post mortem contexts. *Investigative Genetics.* **4**, 7 (2013).

31. O'Donoghue, K., Clapham, A., Evershed, R. & Brown, T. A. Remarkable preservation of biomolecules in ancient radish seeds. *Proc. Roy. Soc. Ser. B* **263**, 541–547 (1996).

32. Palmer, S. A., Moore, J. D., Clapham, A. J., Rose, P. & Allaby, R. G. Archaeogenetic Evidence of Ancient Nubian Barley Evolution from Six to Two-Row Indicates Local Adaptation. *Plos One* **4**, e6301 (2009).

33. Palmer, S. A. *et al.* Archaeogenomic evidence of punctuated genome evolution in *Gossypium*. *Mol. Biol. Evol.* **29**, 2031–2038 (2012).

34. Mayer, K. F. X. *et al.* A physical, genetic and functional sequence assembly of the barley genome. *Nature* **491**, 711–716 (2012).

35. Taubenberger, J. K. *et al.* Reconstruction of the 1918 influenza virus: unexpected rewards from the past. *mBio* **3**, e00201–12; DOI:10.1128/mBio.00201-12 (2012).

36. Taubenberger, J. K. *et al.* Characterization of the 1918 influenza virus polymerase genes. *Nature* **437**, 889–893 (2005).

37. Worobey, M. *et al.* Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature* **455**, 661–664 (2008).

38. Zhang, G. *et al.* Evidence of Influenza A Virus RNA in Siberian Lake Ice. *J. Virol.* **80**, 12229–12235 (2006).

39. Worobey, M. Phylogenetic Evidence against Evolutionary Stasis and Natural Abiotic Reservoirs of Influenza A Virus. *J. Virol.* **82**, 3769–3774 (2008).

40. Hofreiter, M. *et al.* DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nuc. Acids Res.* **29**, 4793–4799 (2001).

41. Brotherton, P. *et al.* Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. *Nuc. Acids Res.* **35**, 5717–5728 (2007).

42. Briggs, A. W. *et al.* Patterns of damage in genomic DNA sequences from a Neandertal. *PNAS* **104**, 14616–14621 (2007).

43. Yu, W. & Schuster, W. Evidence for a site-specific cytidine deamination reaction involved in C to U RNA editing of plant mitochondria. *J. Biol. Chem.* **270**, 18227–18233 (1995).

44. Briggs, A. W. *et al.* Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucl. Acids Res.* **38**, e87; DOI: 10.1093/nar/gkp1163 (2010).

45. Gustafson, G. *et al.* Nucleotide sequence of barley stripe mosaic virus RNA alpha: RNA alpha encodes a single polypeptide with homology to corresponding proteins from other viruses. *Virology* **170**, 370–377 (1989).

46. Petty, I. T. *et al.* Infectious barley stripe mosaic virus RNA transcribed in vitro from full-length genomic cDNA clones. *Virology* **171**, 342–349 (1989).

47. Gustafson, G. *et al.* Nucleotide sequence and genetic organization of barley stripe mosaic virus RNA gamma. *Virology* **158**, 394–406 (1987).

48. Edwards, M. C. Mapping of the seed transmission determinants of barley stripe mosaic virus. *Mol. Plant Mic. Interac.* **8**, 906–915 (1995).

49. Lee, M. Y. & Jackson, A. O. Barley stripe mosaic virus strain Norwich. Direct submission to NCBI Genbank, GI: 347582764, 66 and 71 (2011).

50. Sun, X., Qiu, B. & Li, Y. Barley stripe mosaic virus (China strain). Direct submission to NCBI Genbank, GI: 55925814, 55925816 and 55585709 (2004).

51. Heled, J. & Drummond, A. J. Bayesian Inference of Species Trees from Multilocus Data. *Mol. Biol. Evol.* **27**, 570–580 (2010).

52. Stadler, T. *et al.* Estimating the Basic Reproductive Number from Viral Sequence Data. *Mol. Biol. Evol.* **29**, 347–357 (2012).

53. Li, Y. *et al.* On the origin of smallpox: Correlating variola phylogenics with historical smallpox records. *PNAS.* **104**, 15787–15792 (2007).

54. Zohary, D., Hopf, M. & Weiss, E. *Domestication of Plants in the Old World 4th Edition.* Oxford University Press, Oxford. Pp56–58 (2012).

55. Tyerman, C. *God's War: A New History of the Crusades.* Belknap Press, Cambridge, MA. Pp 787 (2006).

56. Cui, Y. *et al.* Historical variations in mutation rate in an epidemic pathogen, Yersinia pestis. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 577–582 (2013).

57. Drummond, A., Pybus, O. G. & Rambaut, A. Inference of Viral Evolutionary Rates from Molecular Sequences. *Adv. Parasit.* **54**, 331–358 (2003).

58. Gilbert, M. T., Bandelt, H. J., Hofreiter, M. & Barnes, I. Assessing ancient DNA studies. *Trends Ecol. Evol.* **20**, 541–544 (2005).

Author contributions

O.S. carried out the research, R.G.A. designed the research project, A.C. and P.R. provided archaeological materials, Y.L. and J.W. provided sequencing facilities, O.S., R.G.A., A.C. and P.R. analyzed the data and O.S. and R.G.A. wrote the manuscript.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Smith, O. *et al.* A complete ancient RNA genome: identification, reconstruction and evolutionary history of archaeological Barley Stripe Mosaic Virus. *Sci. Rep.* **4**, 4003; DOI:10.1038/srep04003 (2014).