NEON USER GUIDE TO TICK-BORNE PATHOGEN STATUS (DP1.10092)

PREPARED BY	ORGANIZATION	DATE
Sarah Elmendorf	DPS	03/06/2018
Kate Thibault	FSU	03/06/2018

CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	06/19/2017	Initial Release

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

1 DESCRIPTION

1.1 Purpose

This document provides an overview of the data included in this NEON Level 1 data product, the quality controlled product generated from raw Level 0 data, and associated metadata. In the NEON data products framework, the raw data collected in the field, for example, the dry weights of litter functional groups from a single collection event are considered the lowest level (Level 0). Raw data that have been quality checked via the steps detailed herein, as well as simple metrics that emerge from the raw data are considered Level 1 data products.

The text herein provides a discussion of measurement theory and implementation, data product provenance, quality assurance and control methods used, and approximations and/or assumptions made during L1 data creation.

1.2 Scope

This document describes the steps needed to generate the L1 data product Tick-borne pathogen status - the presence/absence of select pathogens in collected ticks - and associated metadata from input data. This document also provides details relevant to the publication of the data products via the NEON data portal, with additional detail available in the file, NEON Data Publication Workbook for Tick-borne pathogen status DP1.10092.001) (AD[05]), provided in the download package for this data product.

This document describes the process for ingesting and performing automated quality assurance and control procedures on the data collected in the field pertaining to TOS Protocol and Procedure: Tick and Tick-Borne Pathogen Sampling (AD[07]). The raw data that are processed in this document are detailed in the file, NEON Data Ingest Workbook for Tick Sampling (NEON.DP0.10093) (AD[04]), provided in the download package for this data product. Please note that raw data products (denoted by 'DP0') may not always have the same numbers (e.g., '10033') as the corresponding L1 data product.

2 RELATED DOCUMENTS AND ACRONYMS

2.1 Associated Documents

AD[01]	NEON.DOC.000001	NEON Observatory Design (NOD) Requirements
AD[02]	NEON.DOC.000913	TOS Science Design for Spatial Sampling
AD[03]	NEON.DOC.002652	NEON Level 1, Level 2 and Level 3 Data Products Catalog
AD[04]	NEON.DP0.10093.001 _dataValidation.csv	NEON Data Ingest Workbook for Tick Sampling (NEON.DP0.10093)
AD[05]	NEON.DP1.10092.001 _variables.csv	NEON Data Publication Workbook for Tick-borne pathogen status DP1.10092.001)
AD[06]	NEON.DOC.000911	TOS Science Design for Vectors and Pathogens
AD[07]	NEON.DOC.014045	TOS Protocol and Procedure: Tick and Tick-Borne Pathogen Sampling
AD[08]	NEON.DOC.000913	TOS Science Design for Spatial Sampling
AD[09]	NEON.DOC.000913	TOS Science Design for Plant Diversity
AD[10]	NEON.DOC.000008	NEON Acronym List
AD[11]	NEON.DOC.000243	NEON Glossary of Terms
AD[12]	OS_Generic_Transitions .pdf	NEON Algorithm Theoretical Basis Document: OS Generic Transitions
AD[13]		NEON's Ingest Conversion Language (NICL) specifications

3 DATA PRODUCT DESCRIPTION

Ticks are sampled in the field using drag and/or flag sampling (Milne 1943, Falco and Fish 1992, Ginsberg and Ewing 1989, Rulison et al. 2013). Collected ticks are enumerated by life-stage and taxonomic group. See data product ticks sampled Using drag cloths (NEON.DP1.10092.001) for details on collection and identification of ticks. Following identification, a subset of ~130 individual tick nymphs per site per year are tested the presence/absence of a suite of bacterial and protozoan pathogens. *Ixodes scapularis* and *Ixodes pacificus* nymphs are targeted for testing of *Anaplasma phagocytophilum*, *Babesia microti*, *Borrelia burgdorferi*, *Borrelia miyamotoi*, and other *Borrelia spp.*. Nymphs in the genera *Dermacentor* and *Amblyomma* are targeted for testing of *Rickettsia spp.* and *Francisella tularensis*.

Ticks transmit numerous pathogens of wildlife, livestock, and humans, including the etiological agent of Lyme disease (Borrelia burgdorferi), the most frequently reported vector-borne disease of humans in the United States. Among arthropod vectors, ticks are particularly sensitive to meteorological conditions and associated physiological constraints, making it highly likely that the demography and biogeography of many tick species, and the pathogens they transmit, will be affected by climate change. Further, the multi-host lifecycles of most tick species increase their ecological connectivity and sensitivity to community-level perturbations that may arise from changes in human land- and resource-use practices.

3.1 Spatial Sampling Design

Tick sampling is conducted at regular intervals by NEON field technicians at terrestrial sites. At each site, six plots are selected for sampling, distributed within the sites proportional to permitted sampling area contained within each National Land Cover Database (NLCD) class. Ticks plots are situated such that the edge of each tick plot must be >150m from other NEON plots and infrastructure; tick plot centers are >500m from eachother and no streams bisect tick plot perimeters. To reduce the probability that the sampling activities in other protocols impact tick diversity and distribution (e.g., technicians inadvertently attracting or redistributing ticks), tick plots centers are offset from distribued plots center according to a specified distance (150m +/- 15m) in a randomly chosen direction established prior to establishment of plots in the field. Sampling occurs by drag or flag sampling the 160m perimeter of each of the six 40m x 40m tick plots (Figure 1). See TOS Science Design for Vectors and Pathogens (AD[06]), TOS Protocol and Procedure: Tick and Tick-Borne Pathogen Sampling (AD[07]), TOS Science Design for Spatial Sampling (AD[08]) and TOS Science Design for Plant Diversity (AD[09]) for further details.

An exception to the design above occurs at NEON's Guanica Forest site (GUAN). At GUAN, the density of vegetation precluded sampling in any of the randomly selected plots using the above design. Instead, field staff located and marked six 'transect paths' where sampling is feasible. Each transect path contains between 80m^2 and 160m^2 of sampled area. Dragging and/or flagging then occurs over one more 10m long segments within each plot, with each segments generally within of another segment. Where possible, transect paths are located within 300m from the center of a mammal plot; 300m from the center of a bird plot; and 300m from the center of a Distributed Base Plot, with preference given to co-location with mammal plots where all three co-location criteria cannot be met.

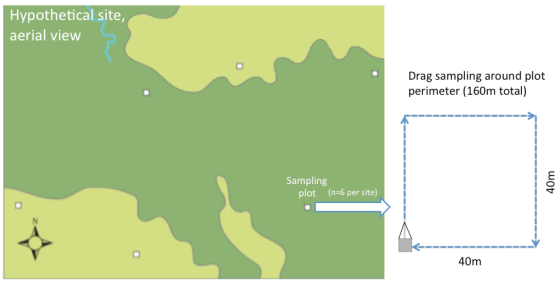

Figure 1: Drag and/or flag sampling occurs around the 160m perimeter of a 40mx40m plot

3.2 Temporal Sampling Design

Ticks are sampled during the growing season (determined by site-specific average greening period). During the growing season, sampling is conducted every three weeks at sites where ticks have previously been detected and every six weeks elsewhere. Sampling also occurs only if the high temperature on two consecutive days prior to planned sampling exceeds 0°C. For additional details, see the TOS Protocol and Procedure: Tick and Tick-Borne Pathogen Sampling.

3.3 Variables Reported

All variables reported from the field or laboratory technician (L0 data) are listed in the file, NEON Data Ingest Workbook for Tick Sampling (NEON.DP0.10093) (AD[04]). All variables reported in the published data (L1 data) are also provided separately in the file, NEON Data Publication Workbook for Tick-borne pathogen status DP1.10092.001) (AD[05]).

Field names have been standardized with Darwin Core terms (http://rs.tdwg.org/dwc/; accessed 16 February 2014), the Global Biodiversity Information Facility vocabularies (http://rs.gbif.org/vocabulary/gbif/; accessed 16 February 2014), the VegCore data dictionary (https://projects.nceas.ucsb.edu/nceas/projects/bien/wiki/VegCore; accessed 16 February 2014), where applicable. NEON TOS spatial data employs the World Geodetic System 1984 (WGS84) for its fundamental reference datum and GEOID09 for its reference gravitational ellipsoid. Latitudes and longitudes are denoted in decimal notation to six decimal places, with longitudes indicated as negative west of the Greenwich meridian.

Some variables described in this document may be for NEON internal use only and will not appear in downloaded data.

3.4 Spatial Resolution and Extent

The finest resolution at which spatial data are reported is a single plot.

plotID (ID of plot within site) → **siteID** (ID of NEON site) → **domainID** (ID of a NEON domain).

The basic spatial data included in the data downloaded include the latitude, longitude, and elevation of the centroid of the plot where sampling occurred + associated uncertainty due to GPS error and plot width. Shape-

files of all NEON Terrestrial Observation System sampling locations can be found in the Document Library: http://data.neonscience.org/documents.

3.5 Temporal Resolution and Extent

The finest resolution at which temporal data are reported is the range between the startDate and endDate of each tick collection event (the duration during which dragging and/or flagging a single plot occurred).

The NEON Data Portal provides data in monthly files for query and download efficiency. Queries including any part of a month will return data from the entire month. All queries, regardless of the date range specified, will include a copy of tck_pathogenqa which provides data from the pathogen lab about analytical quality control. Code to stack files across months is available here: https://github.com/NEONScience/NEON-utilities

3.6 Associated Data Streams

subsampleID is a linking variable that ties specific tick identifications and associated metadata to the Ticks sampled using drag cloths data product (DP1.10093). Visual assessments of the presence of ectoparasites, by life stage, on captured mammals can be found in the Small mammal box trapping data product (DP1.10072).

3.7 Product Instances

There are a maximum ~130 ticks per site per year submitted for pathogen testing. Each tick may be tested for multiple pathogens, yielding a variable number of tck_pathogen records per site per year.

3.8 Data Relationships

The protocol dictates that each plot is sampled once per sampling bout (six expected records per eventID in the Ticks sampled using drag cloth product tck_fielddata). A record from Ticks sampled using drag cloth product tck_fielddata may have zero (if no ticks present) or more child records in Ticks sampled using drag cloth product tck_taxonomyProcessed. An equivalent number of records will occur in Ticks sampled using drag cloth product tck_taxonomyRaw. A record from tck_taxonomyProcessed may have zero (if the pertinent subsampleIDs are not sent for pathogen testing) or more records in the tck_pathogen table. Duplicates and/or missing data may exist where protocol and/or data entry aberrations have occurred; users should check data carefully for anomalies before joining tables.

tck_fielddata.csv (NEON.DP1.10093) - > One record expected per plotID for each eventID, generates a single sampleID if ticks are present

tck_taxonomyProcessed.csv (NEON.DP1.10093) - > Multiple record expected per sampleID, generates a unique subsampleID for each species*life stage combination found per sampleID. The taxonomic nomenclature in this file has been standardized and desynonymized according to NEON's master taxonomy for ticks.

tck_taxonomyRaw.csv (NEON.DP1.10093 - expanded package) - > Multiple record expected per sampleID, generates a unique subsampleID for each species*life stage combination found per sampleID. The taxonomic nomen-

clature in this file reflects the verbatim identifications provided by the external taxonomist and may contain synonyms.

tck_pathogen.csv (NEON.DP1.10092) - > Multiple record expected per subsampleID, generates a unique testingID for each tick selected for pathogen testing. Note not all subsampleIDs contribute to testing; some are simply archived in 95% EtOH.

tck_pathogenqa.csv (NEON.DP1.10092 - expanded) - > One record expected per batchID, contains relevant batch-level quality assurance procedures for each batch tested over all dates testingID for each tick selected for pathogen testing. Note not all subsampleIDs contribute to testing; some are simply archived in 95% EtOH.

4 DATA QUALITY

4.1 Data Entry Constraint and Validation

Constraints and data validation are implemented during the process of ingest into the NEON database. The product-specific data constraint and validation requirements built into data entry applications and database ingest are described in the document NEON Data Ingest Workbook for Tick Sampling (NEON.DP0.10093), provided with every download of this data product. Contained within this file is a field named 'entryValidationRulesParser', which describes syntactically the validation rules for each field built into the data entry application. Data entry constraints are described in Nicl syntax in the validation file provided with every data download, and the Nicl language is described in NEON's Ingest Conversion Language (NICL) specifications ([AD[13]).

4.2 Automated Data Processing Steps

Following data entry into a mobile application of web user interface, the steps used to process the data through to publication on the NEON Data Portal are detailed in the NEON Algorithm Theoretical Basis Document: OS Generic Transitions (AD[12]).

4.3 Data Revision

All data are provisional until a numbered version is released; the first release of a static version of NEON data, annotated with a globally unique identifier, is planned to take place in 2020. During the provisional period, QA/QC is an active process, as opposed to a discrete activity performed once, and records are updated on a rolling basis as a result of scheduled tests or feedback from data users. The Change Log section of the data product readme, provided with every data download, contains a history of major known errors and revisions.

4.4 Quality Flagging

The **dataQF** field in each data record is a quality flag for known errors applying to the record. Please see below for an explanation of **dataQF** codes specific to this product.

fieldName	value	definition
dataQF	legacyData	Data recorded using a paper-based workflow that did not implement the full suite of quality control features associated with the interactive digital workflow

4.5 Analytical Facility Data Quality

Batches of samples analyzed for pathogens are checked for contamination via a negative control on nucleic acid extraction (tck_pathogenqa_pub.controlType='extractionNegative'), a negative control for master mix (tck_pathogenqa_pub.controlType='mixedSampleNegative'), a negative control for master mix (tck_pathogenqa_pub.controlType='masterMixNegative'), and a known positive control (tck_pathogenqa_pub.controlType='positive'). Details on these control measures can be found in the external lab SOP.

5 REFERENCES

Falco, R. C., and D. Fish. 1992. A comparison of methods for sampling the deer tick, Ixodes dammini, in a Lyme disease endemic area. Experimental & Applied Acarology 14:165–173.

Ginsberg, H. S., and C. P. Ewing. 1989. Comparison of flagging, walking, trapping, and collecting from hosts as sampling methods for northern deer ticks, Ixodes dammini, and lone star ticks, Amblyomma americanum (Acari, Ixodidae). Experimental and Applied Acarology 7:313–322.

Milne, A. 1943. The comparison of sheep tick populations (Ixodes ricinus L). Annals of Applied Biology 30:240–250.

Rulison, E. L., I. Kuczaj, G. Pang, G. J. Hickling, J. I. Tsao, and H. S. Ginsberg. 2013. Flagging versus dragging as sampling methods for nymphal Ixodes scapularis (Acari: Ixodidae). Journal of Vector Ecology 38:163–167.