# Supplementary Data

## Section S1: Additional information about cliff-face seep definition

Our definition of cliff-face seeps permitted slight spatial differences between cliff faces and groundwater discharge locations: habitat features satisfying the above criteria were considered valid cliff face seeps if they were located within ~200m downslope of visually estimated dry-season groundwater emergence points; this distance was subjectively chosen based on related patterns of amphibian habitat use. Dry-season emergence points were estimated based on the most upslope location that had water-stained rocks or contained plant species (e.g., hydrophilic mosses and algae) dependent on madicolous habitats.

## Section S2: Methods for obtaining downscaled climate data

We initially hypothesized that downscaled climate data might be more useful than coarser-scale climate data for predicting locations of seeps with small contributing aquifers, making it necessary to test climate-related predictor variables across multiple resolutions. Therefore, we used a kriging approach to downscale each of 19 available ~800x800m resolution BioClim variables (Fick and Hijmans 2017) to a ~180x180m resolution within the "KrigR" package in R (Kusch and Davy 2022). We used elevation as a downscaling covariate for all variables, although we acknowledge limitations of this approach for precipitation-related climate variables (Jiang et al. 2018). Prior to downscaling, we calculated BioClim variables using the "Dismo" package in R (Hijmans et al. 2017) and historical (1991-2020) 800m-scale PRISM precipitation and temperature data (Daly and Bryant 2013). We used PRISM data instead of more expansive WorldClim data available, because WorldClim was largely inaccurate at depicting precipitation patterns across portions of the PNW (e.g., the Olympic Peninsula and Idaho Rockies) in comparison to PRISM (pers. obs.). However, because 800m-scale historical PRISM climate data (1991-2020) were not available for Canada, we took a two-step approach to estimate Canadian climate data. First, we modeled bioclimatic variables derived from PRISM climate data across the PNW United States (Washington, Oregon, Idaho, Northern California, and Western Montana) using GLMs (McCullagh 2019), with ~1km WorldClim data as the predictor variable for each climate metric and corresponding ~800m PRISM data as the response variable. We then used these models to predict 800m-scale historical (1991-2020) PRISM climate data for Canada.

For both our ~800x800m and downscaled ~180x180m modeling approach, we constructed initial BRTs (i.e., before removing uninformative and redundant variables) including

the same additional 31 climate, landscape, and region-related predictor variables as previously (Table S1). As downscaled data were computationally intensive to produce, we only downscaled areas located <1km away from BRT presence and absence points in this comparative exercise. We refined BRTs using the same methods as previously, then generated and logit-transformed predictions of the input data and assessed their similarity based on Pearson's correlation coefficient (Sachs 2012). We determined whether downscaling BioClim variables was necessary partly based on uniqueness (or lack thereof; based on correlation) of the predictions, and also by comparing model performance based on cross-validated AUC scores using $k = 10$ folds (Hastie et al. 2009).

## Section S3: Methods for logger data

We took multiple steps to clean temperature logger data to ensure that they were relevant for inferring refugial capacity. First, we removed days with incomplete logger data or snow cover and calculated daily temperature means, standard deviations, and ranges for each logger for the remaining days. Temperature data from water-based loggers were only valid (i.e., measuring water rather than air temperature) if seeps were not dry, and measurements from these loggers converged with air temperatures after seeps dried. Therefore, to determine which water-based logger data was invalid due to dry seeps, we used a generalized linear mixed model (GLMM) (Bolker et al. 2009) to predict the probability that a given logger was logging air versus water temperatures on a given day, using the daily mean, standard deviation, and range for temperature, two-way interactions (temperature range*standard deviation and temperature mean*range), and a random effect of region (Olympics, Rainier, or North Fork Clearwater) as predictor variables. We fit this model using a subset of data in which we knew with certainty whether loggers were monitoring water or air temperatures. Based on visual observations and snowmelt trends, all seeps with temperature loggers contained flowing water prior to July 7, 2022 (or possibly slightly later). Therefore, to model whether a temperature logger was measuring air or water temperature, we discarded uncertain water logger data from after July 7, 2022, coded water logger date before this data as 1s, and coded all air-based logger data as 0s regardless of date. We used these data to build GLMMs and then generate predictions of whether water-based loggers were logging air temperatures (i.e., dry seeps) or water temperatures across our entire dataset. We discarded all water-based logger data with a predicted probability of seep wetness (i.e., logging of water rather than air temperatures) below 0.5%, as this threshold yielded

the greatest accuracy for classifying seeps as wet or dry based on a comparison with visual observations.

We calculated "relative thermal variation" based on the average daily difference between a logger's observed temperature range and the observed temperature ranges of all corresponding loggers at other seeps. This served to account for asymmetrical expected temperature variation among loggers due to differential timing of installation. For both air and water temperature loggers, we calculated the $90^{th}$ and $10^{th}$ percentiles for daily mean temperatures at the logger and subtracted the former from the latter to estimate "observed thermal variation". We then pooled daily temperature means across all loggers (but not days) within each logger type (air or water) and calculated "expected thermal variation" for each day and logger type by subtracting the $10^{th}$ percentile from the $90^{th}$ percentile of daily values for these pooled logger data. We then subtracted this "expected thermal variation" for each day from the "observed thermal variation" for each focal logger and calculated the mean across all days to generate the "relative thermal variation" for each logger type at each seep. All days in which <10 loggers of the same type (air or water) were logging valid data simultaneously were excluded from this calculation due to a limited sample size for comparing loggers to one another. We used only the middle 80% of daily temperature data—and not brief extremes beyond this range—because most seep-associated species can temporarily burrow to avoid briefly extreme or volatile surface temperatures. In contrast, more sustained variability (e.g., within this middle 80%) could make surficial and near-surface microhabitats more broadly unsuitable for seep-dwellers, due to less predictably tolerable surface temperatures.

## Section S4: Generating output maps

We used the "raster" function in R to generate 30m-resolution output maps with predictions of seep occurrence probabilities—as well as seep thermal and hydrologic properties—estimated for each map pixel across our study region. In addition, we generated a single map summarizing "overall seep refugial potential" using the following formulas:

$$RCS = SPP - RATV - RWTV$$

$$OSRP = (RCS - minimum(RCS)) * SOC$$

where RCS = refugial capacity score, SPP = seep permanence probability, RATV = relative air temperature variation, RWTV = relative water temperature variation, OSRP = overall seep refugial potential, and SOC = seep occurrence probability. We z-scored SPP, RATV, and RWTV prior to the above calculations; thus, the above formulas assign equal weight to the three parameters associated with seep thermal and hydrologic stability and condition their cumulative importance on the probability that a seep occurs within the pixel. We used z-scores in this equation to ensure that our air and water temperature metrics were on the same scale and could thus be weighted equally in our equation by simply adding them together. Prior to z-scoring, RATV and RWTV were multiplied by -1 to reverse their ordering, such that a z-score reflected relative temperature stability (and thus refugial potential) rather than variation. To preserve zero as a meaningful baseline value for refugial capacity, minimum RCS was subtracted from each pixel's RCS value prior to multiplying by SOC. We clipped all output maps to areas where we deemed seep predictions to have "high" or "medium" confidence (Appendix 4).

## Section S5: Defining confidence levels for outputs

We identified areas of high, medium, low, or very low confidence for our cumulative seep predictions based on values of important predictor variables in each pixel within the PNW (including Washington, Oregon, Idaho, northern California, southern British Columbia, and the Rocky Mountains within Montana and southwest Alberta). Boundaries of these four confidence levels were necessarily arbitrary but were nonetheless necessary to designate, to acknowledge the possibility that environmental drivers of seep refugia may vary for climates and subregions outside the scope of our sampling efforts (e.g., semi-arid portions of the interior PNW). We considered a pixel's predictions to be "high confidence" if values for all predictor variables within our final seep distribution model fell between the $2.5^{th}$ and $97.5^{th}$ percentiles (i.e., middle 95%) of values for model training points, and if the pixel was additionally located within one of our sampling regions (i.e., the Olympic Peninsula, Cascades Range between Mount Rainier and the Columbia River, Clearwater River Drainage, or Coeur d'Alene River Drainage). "Medium confidence" pixels included (1) those located within sampling regions and featuring predictor variables within the overall range for training data but a single variable falling outside of the middle 95% of training values, and (2) those located outside of sampling regions but with values falling within the middle 95% of training data for every predictor variable. "Low confidence" pixels included (1) those located within sampling regions and with all predictor variables falling within the overall range for training data but multiple predictor variables falling outside of the middle 95% of training values, and (2) those located outside of sampling regions and with all predictor variables falling within the overall range for training data but a single predictor variable falling outside of the middle 95% of training values. We classified seep predictions as "very low confidence" for all other map pixels due to unrepresented regions and/or predictor variable data.

Pixels located in Canada and California were automatically considered “very low confidence”, due to a lack of available riparian climate corridor index data (Krosby et al. 2018). Predictions from pixels with “low” or “very low” confidence levels are excluded from primary figures herein to avoid overgeneralization of our results. However, we chose to include these high-uncertainty predictions as supplementary files so that they are usable as preliminary hypotheses of potential seep locations in future studies.

## Supplementary Tables and Figures

**Table S1.** List of variables used to fit initial seep distribution models.

| Variable | Units | Resolution | Data Source |
|---|---|---|---|
| Elevation | Meters | 30m | USGS 3DEP |
| Slope | Degrees | 30m | USGS 3DEP |
| Aspect difference from north | Degrees | 30m | USGS 3DEP |
| Average slope within 300m | Degrees | 30m | USGS 3DEP |
| Standard deviation in aspect distance from north within 30m | Degrees | 30m | USGS 3DEP |
| Topographic Wetness Index (TWI) | Index | 30m | USGS 3DEP |
| Maximum TWI within 30m | Index | 30m | USGS 3DEP |
| Flow accumulation | Index | 30m | USGS 3DEP |
| Maximum flow accumulation within 30m | Index | 30m | USGS 3DEP |
| Curvature | Degrees/Meter | 30m | USGS 3DEP |
| Vertical distance below highest point within 1km | Meters | 30m | USGS 3DEP |
| Distance to NHD flowline | Meters | ~10m | National Hydrography Dataset (NHD) |
| Tree canopy | Percent | 30m | National Land Cover Dataset (NLCD) 2016 |
| Average stand height | Meters | 30m | Landscape Ecology Modeling, Mapping, and Analysis (LEMMA) Team |

| Variable | Units | Resolution | Data Source |
|---|---|---|---|
| Volume of live trees ("VPH") | $m^3$/ha | 30m | Landscape Ecology Modeling, Mapping, and Analysis (LEMMA) Team |
| Landsat (Aqua Satellite) Enhanced Vegetation Index (extracted from September data with < 5% cloud cover) | Index | 250m | NASA Moderate Resolution Imaging Spectroradiometer (MODIS) |
| Dominant rock type | Category | N/A (extracted from shapefile at 30m resolution) | Various |
| Landsat Band 7 surface reflectance | Index | 30m | NASA Harmonized Landsat (HLS) Sentinel-2 |
| Region | N/A (categorical) | N/A | N/A |
| Base-Flow Index | Index | 1km | Wolock (2003) |
| Soil water storage | Preceding 10-yr mean of mean daily soil water storage (in mm) | 1km | Tercek (2018) |
| Snow-water equivalent | Preceding 10-yr mean of mean cumulative sum of daily snow-water equivalent (in $kg/m^2$) during winter | 1km | DAYMET (Thornton et al. 2020) |
| Riparian climate corridor index | Index values (0 – 1) attributed to points at the headwaters associated with each potential riparian corridor (all non-corridors coded as 0) | 90m | Krosby et al. (2018) |

| Variable | Units | Resolution | Data Source |
|---|---|---|---|
| Maximum riparian climate corridor index within 90m | Index | 90m | Krosby et al. (2018) |
| Bioclimatic variables ($n$ = 19) | Various | 180m & 1km | PRISM |
| Near-surface relative humidity | Percentage | 1km | CHELSA-BIOCLIM+ (Brun et al. 2022) |
| Vapor pressure deficit | Kilopascals | 1km | CHELSA-BIOCLIM+ (Brun et al. 2022) |
| Potential evapotranspiration | Millimeters | 1km | CHELSA-BIOCLIM+ (Brun et al. 2022) |
| Climate moisture index | Index | 1km | CHELSA-BIOCLIM+ (Brun et al. 2022) |
| Growing season length | Days | 1km | CHELSA-BIOCLIM+ (Brun et al. 2022) |
| Growing season temperature | Celsius | 1km | CHELSA-BIOCLIM+ (Brun et al. 2022) |
| Growing season precipitation | Millimeters | 1km | CHELSA-BIOCLIM+ (Brun et al. 2022) |

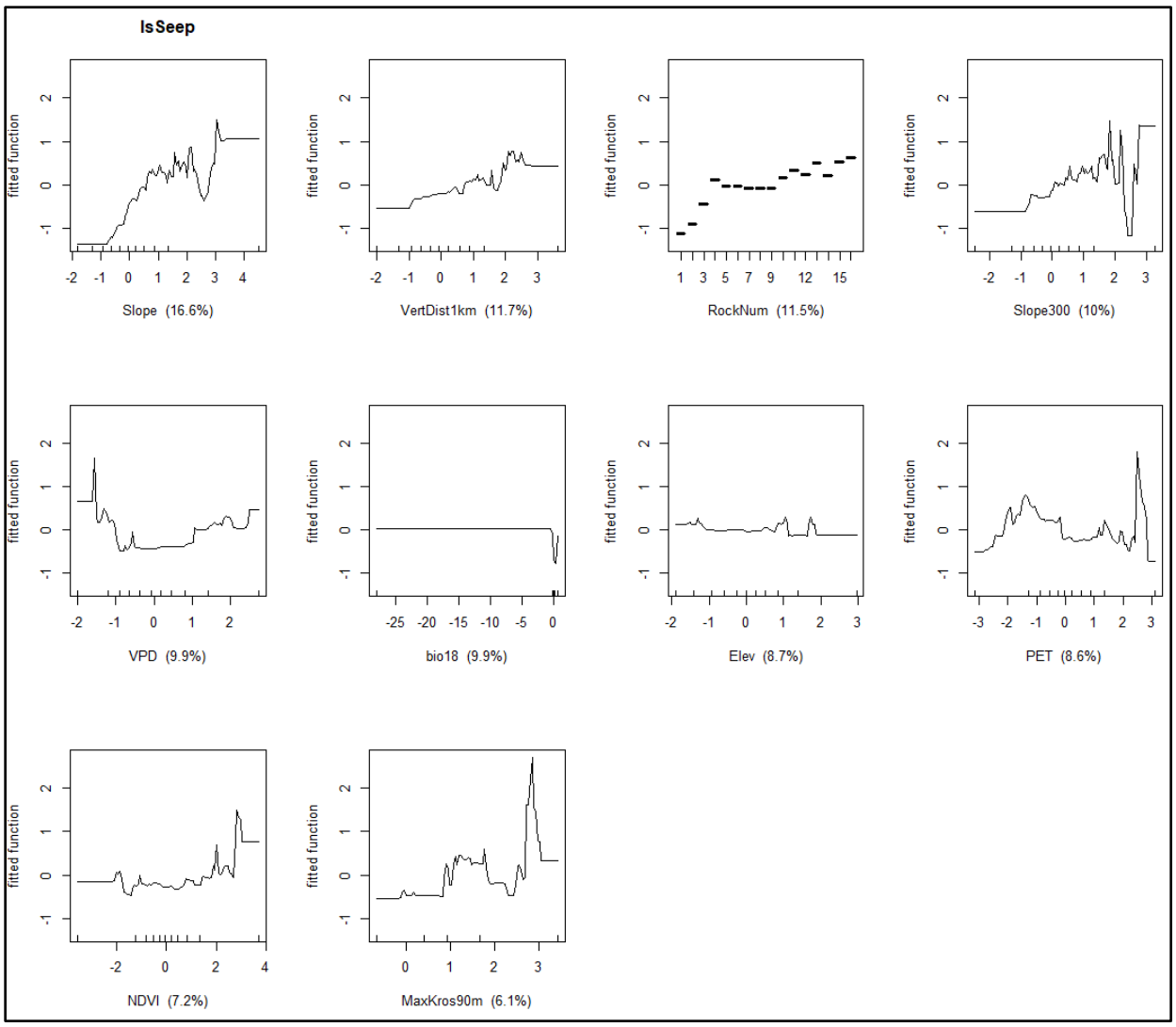


**Figure S1.** Response curves for variables included in the final seep distribution model. Slope300 = average slope within 300m; VertDist1km = vertical distance between pixel and highest point within 1km; Slope = slope; RockNum = dominant rock type; bio18 = precipitation of the warmest quarter; VPD = average vapor pressure deficit; PET = average potential evapotranspiration; Elev = elevation; SWE10yr = average snow water equivalent over past 10 years; NDVI = normalized difference vegetation index.

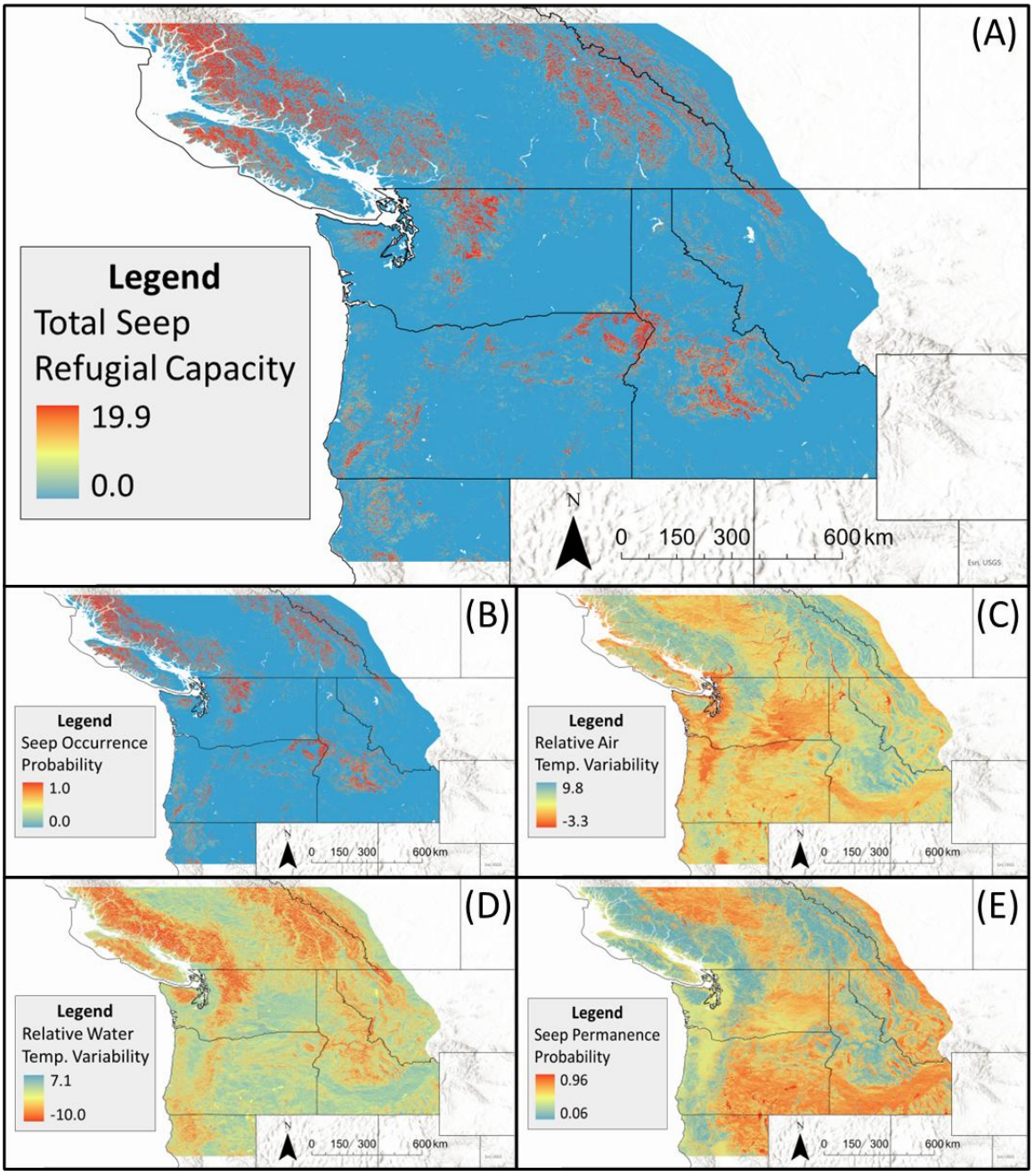


**Figure S2.** Seep predictions for the entire study area, including (A) overall seep refugial capacity, (B) probability of seep occurrence, (C) air temperature stability, (D) water temperature stability, and (E) seep permanence. Predictions for areas with low confidence (Figure S3) should be treated with caution.

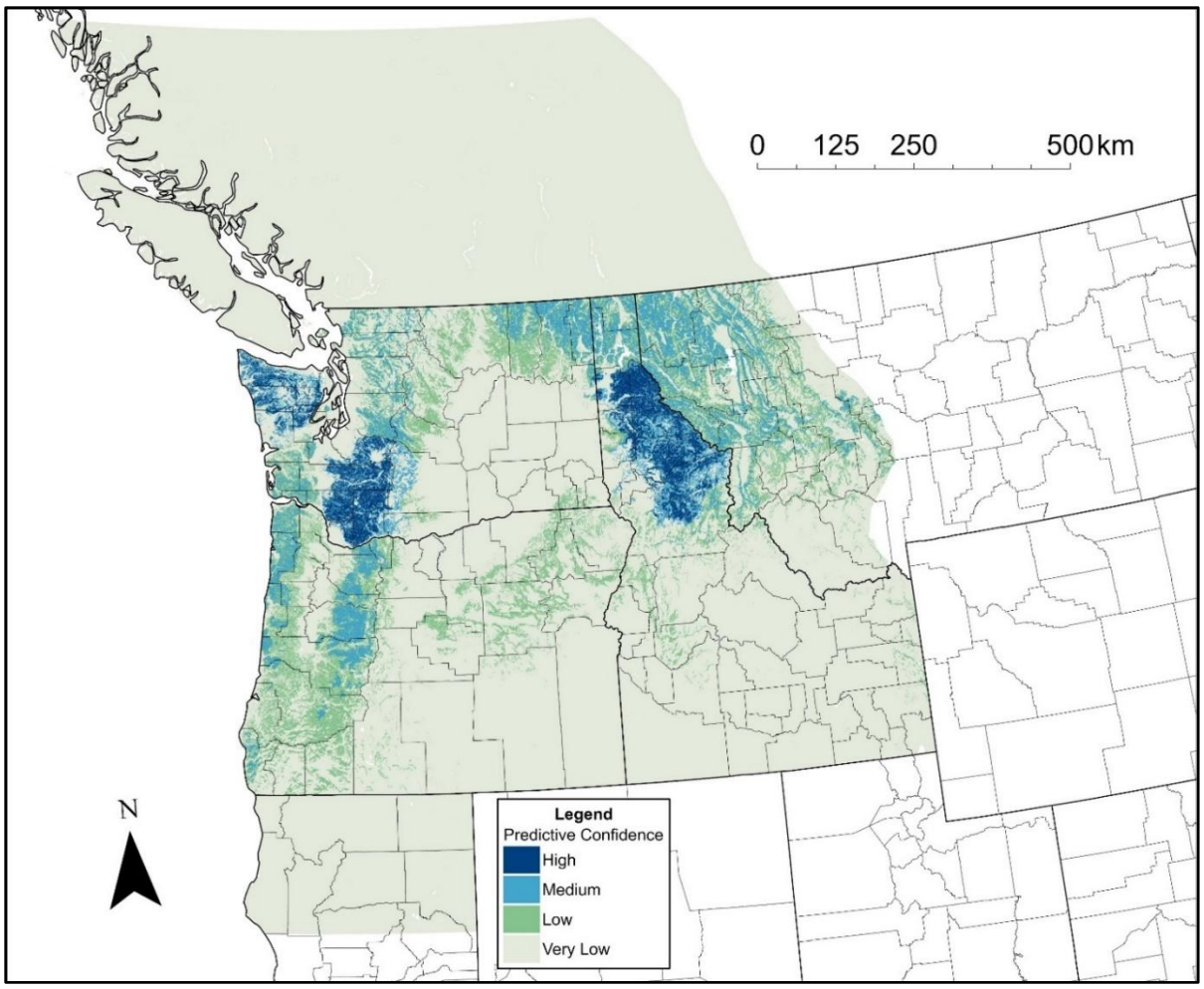


**Figure S3.** Relative confidence levels in seep predictions, based on geographic and climatic overlaps with sampled areas and availability of informative predictor variables.