ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

Hydrocarbon Generation from Low-Mature Saline Lacustrine Sediments Studied Using Machine Learning and Chemometric Methods: The Succession of the Sikeshu Sag, Junggar Basin, NW China

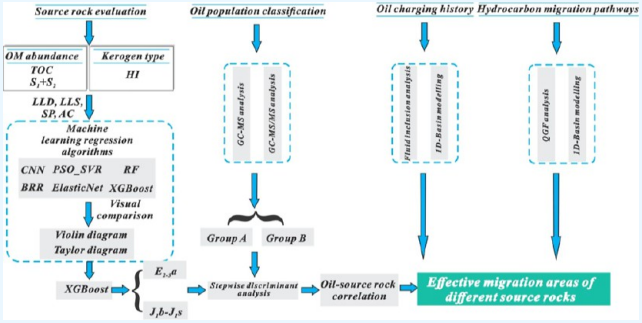
Tengfei Yu,* Zhiping Wu, Ruichao Guo, Guanlong Zhang, Yuejing Zhang, Fengkai Shang, and Lin Chen

Cite This: *ACS Omega* 2023, 8, 10314−10334

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Significant attention has been given to the extensive development of saline environments in petroliferous basins. Further exploration and studies have discovered that saline environments, such as those for the deposition of source rocks in the Paleogene Anjihaihe ($E_{2\text{-}3}a$) Formation of the Sikeshu Sag, are ubiquitous in terrestrial lake basins. Previous studies have suggested that the oil reservoirs in the Sikeshu Sag and its peripheral regions are predominantly derived from the black mudstone and coal measures of the Lower Jurassic Badaowan (J_1b) Formation. However, with deeper exploration of the study area, a growing number of reservoirs with geochemical characteristics different from the J_1b oil source have been discovered, indicating that there are oil sources other than the J_1b source rocks. In this study, various machine learning algorithms were used (random forest, RF; convolutional neural networks, CNN; extreme gradient boosting, XGBoost; ElasticNetCV; Bayesian Ridge; and particle swarm optimization-support vector regression) to select the most suitable algorithm for predicting and comparing the quality of potential source rocks. A violin plot and Taylor diagram were applied to visually compare the reliability and application effectiveness of the models. The results demonstrated that XGBoost and RF can become essential tools for predicting the quality of potential source rocks. Moreover, the measured and predicted values of total organic carbon (TOC), hydrocarbon potential ($S_1 + S_2$), and hydrogen index indicate that there are three main source rocks: the $E_{2\text{-}3}a$, Lower Jurassic Sangonghe (J_1s), and J_1b formations. The thermal maturity of the $E_{2\text{-}3}a$ source rocks is still early mature because of the saline−brackish water nature of these rocks, although large-scale hydrocarbon generation and expulsion can be achieved in the early mature stage. Based on their geochemical characteristics and stepwise discriminant analysis, the oils in the Sikeshu Sag and its peripheral regions can be categorized into two types: groups A and B. Comprehensive organic geochemical evidence suggests that genetically, group A oils are originated from $E_{2\text{-}3}a$ less-mature saline lacustrine sedimentary rocks, while group B oils indicate similar affinity to the Jurassic source. Fluid inclusion microthermometry and one-dimensional basin modeling showed that the oil charging periods of group A and B oils were Middle-Late Miocene (13−8 Ma) and Late Oligocene (23−20 Ma), respectively. Quantitative grain fluorescence (QGF) analysis further propose that the hydrocarbon supply region of the $E_{2\text{-}3}a$ sources is mainly located east of the Western Chepaizi Uplift and the interior area of the Sikeshu Sag, which breaks through the previous understanding that the Jurassic coal-derived oil source is the only main contributor in this study area. The research results can be widely applied to assess the petroleum resources of source rocks in similar areas worldwide.



1. INTRODUCTION

Globally, changes in climate circa 2050 have a moderate impact on energy consumption of 7−17% depending on the degree of warming. Almost all continents experience increases in energy demand, driven by the commercial and industrial sectors.[1,2] An increasing number of saline−alkali lacustrine source rocks have been discovered worldwide, including the Green River Piceance Creek Basin, Uinta Basin, Gulf of Mexico Basin in the United States; the Campos Basin in Brazil; the Red Sea Basin in Europe; and the Jianghan Basin, Junggar Basin, and Qaidam Basin in China,[3−9] exhibiting great potential for petroleum exploration in saline lacustrine basins. Therefore, it is important to discuss oil

Received: December 15, 2022
Accepted: February 20, 2023
Published: March 7, 2023

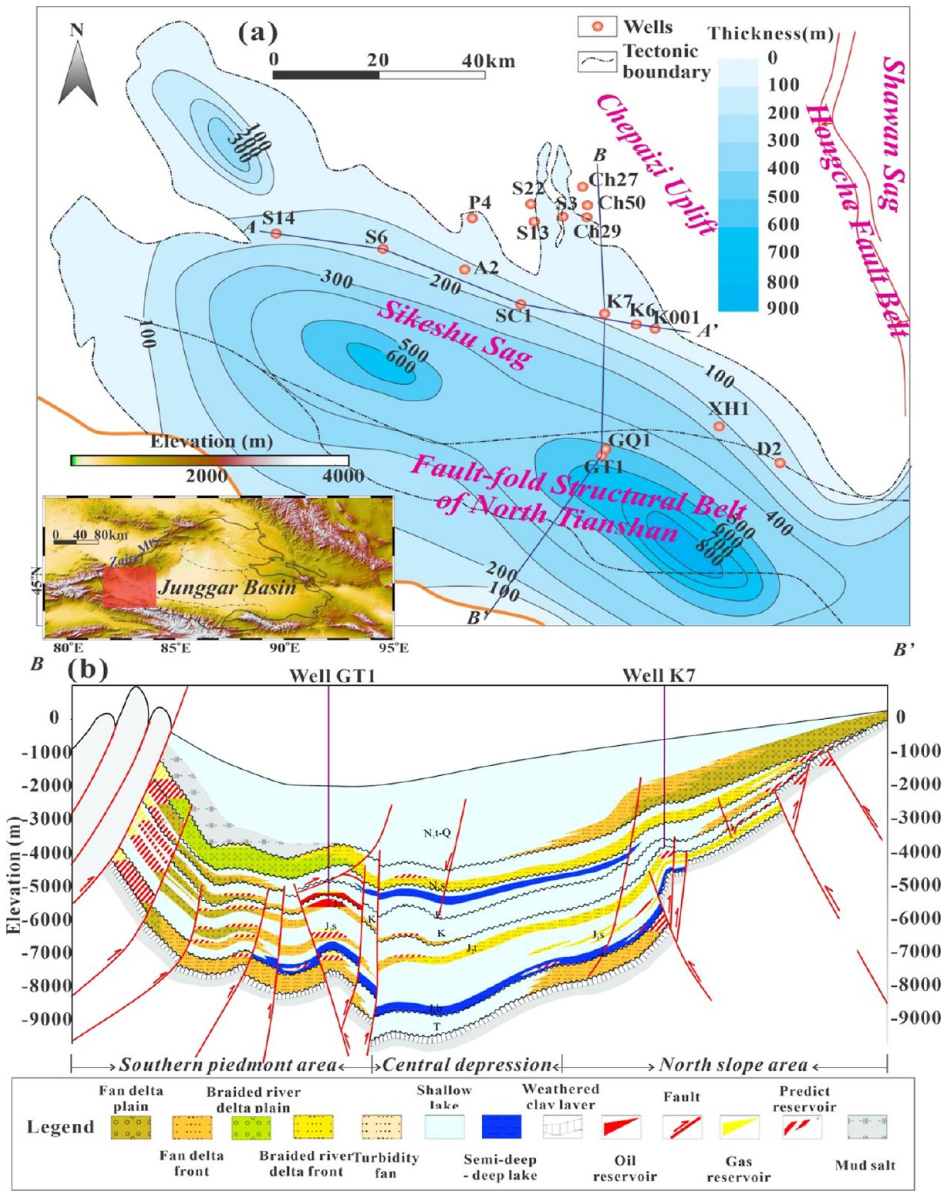


Figure 1. (**a**) Geological map outlining the regional tectonic elements of the Sikeshu Sag and (**b**) *NS*-trending profile showing its stratigraphic framework.

and gas development in saline lake environments in the context of petroleum geology and geochemistry.

Recently, significant attention has been given to insights into the potential oil source of the Sikeshu Sag, other than the Lower Jurassic Badaowan (J_1b) and Sangonghe (J_1s) formations, that is, Triassic and Cretaceous intervals, and the Paleogene Anjihaihe ($E_{2\text{-}3}a$) Formation.[10−18] The discovery of several hydrocarbon-bearing tectonic belts, that is, the Kayindike, Dushanzi, and Gaoquan anticlines, indicates that the Sikeshu Sag has good oil and gas exploration potential. In 2019, a major exploration breakthrough occurred in the Sikeshu Sag, when a high-yield light oil flow with a daily production of >1200 m^3/day and a low gas-to-oil ratio of 265.2 was discovered in the "bottom sandstone" of K_1q.[19−22] To date, little consensus has been reached on the potential oil source of saline lacustrine sediments in $E_{2\text{-}3}a$ in the early-mature stage, and the hydrocarbon migration pathway and accumulation model of the Sikeshu Sag remain unclear.[23−30]

Mainstream thinking suggests that the oil reservoirs in the Sikeshu Sag and surrounding areas are mainly derived from the coal measure and black mudstones of J_1b, which migrate through faults and sheet sand bodies.[31−33] However, the detailed oil−source rock correlation demonstrates that a considerable amount of crude oil exhibits dissimilar genetic affinities with the coal measure and black mudstones of J_1b. Hence, whether multiple sets of effective source rocks are developed, their hydrocarbon generation potential and migration of hydrocarbons based on multiple oil sources are key issues to solve in the next step of identifying favorable exploration targets in the Sikeshu Sag. Therefore, in this study, numerous machine learning (ML) algorithms, including random forest (RF), convolutional neural network (CNN), support vector regression (SVR), extreme gradient boosting (XGBoost), ElasticNetCV, Bayesian Ridge, and particle swarm optimization-SVR (PSO-SVR), were executed to select the most suitable algorithm for predicting and comparing the source rock quality of $E_{2\text{-}3}a$ and J_1b.

https://doi.org/10.1021/acsomega.2c07991
ACS Omega 2023, 8, 10314−10334

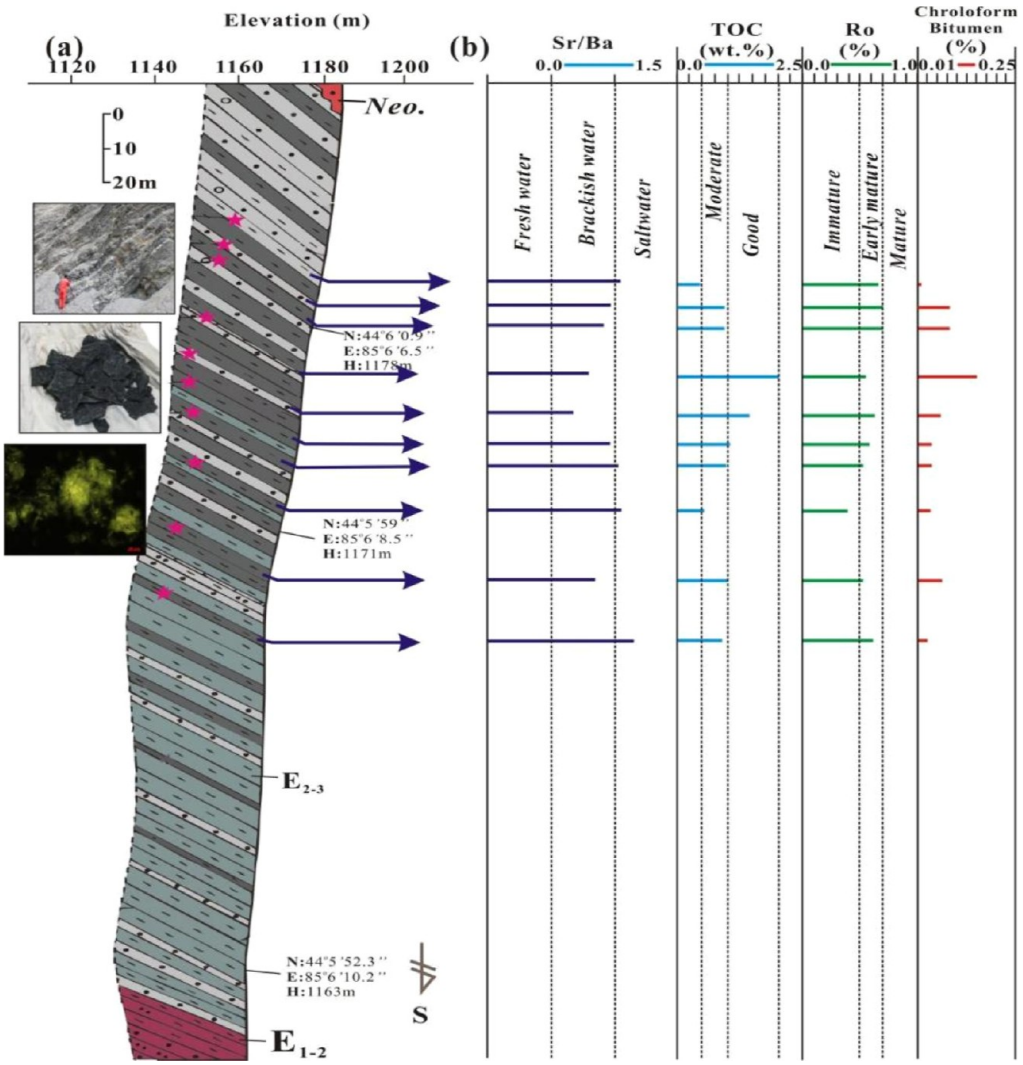


Figure 2. (**a**) Stratigraphic chart and (**b**) geochemical profiles of $E_{2\text{-}3}a$ from the Anjihaihe outcrop showing the stratigraphic characteristics and paleoenvironment of $E_{2\text{-}3}a$.

Extensive literature implies that the brackish water to saltwater transition with appropriate salinity (Cl^-) exhibits a distinct enhancement in the formation of low-maturity oil during the early thermal stage.[34−40] Furthermore, the saline−alkali environment is conducive to the prosperity of algae, that is, laminae and dinoflagellates, and the water stratification of salt lakes provides good preservation conditions for high-quality source rocks and accelerates the transformation of organic matter (OM) to hydrocarbons in saline lacustrine sediments.[41−48] A comprehensive investigation of the hydrocarbon generation potential of Paleogene less-mature saline lacustrine sediments and the reconstruction of oil charging and migration history are important for the implementation of favorable oil and gas exploration targets and for increasing oil production on a large scale.

2. GEOLOGICAL SETTINGS

2.1. Tectonic Locations. The Sikeshu Sag, which has undergone extensive geophysical investigation, is an inherited low-lying region that developed in the Mesozoic in the NW margin of the Junggar Basin, NW China (Figure 1a). It is bound by the Yilinhebiergen Mountain in the southwest, the Hongche Fault Belt to the east, and the Aika Fault Belt to the north,

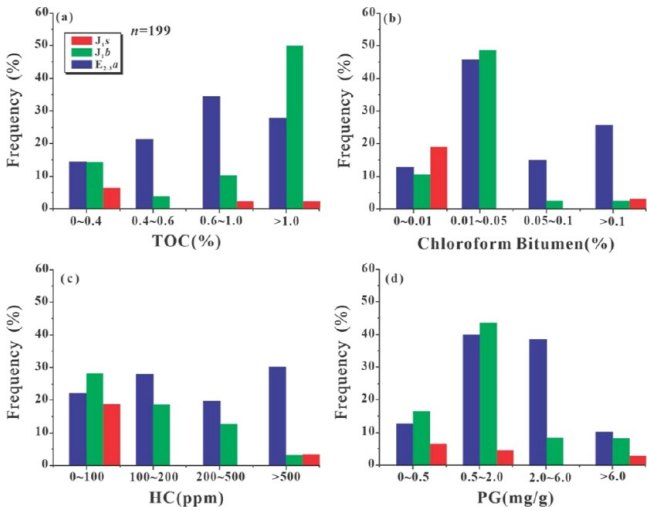


Figure 3. Comprehensive histogram of (**a**) TOC, (**b**) chloroform bitumen "A", (**c**) HC values, and (**d**) PG, exhibiting the differential hydrocarbon generation potential of various sets of source rocks in the Sikeshu Sag.

https://doi.org/10.1021/acsomega.2c07991
ACS Omega 2023, 8, 10314−10334

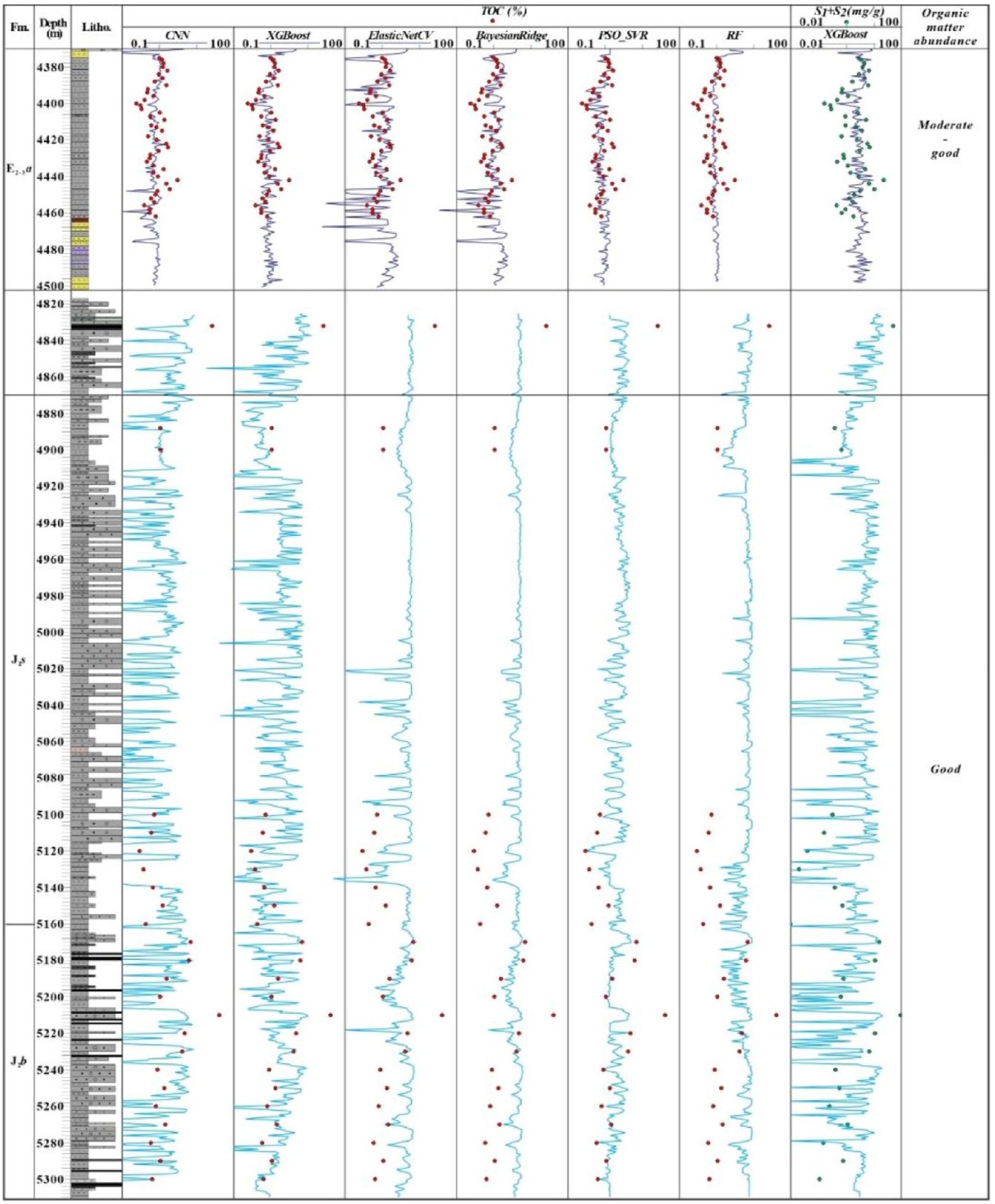


Figure 4. Comprehensive histogram of predicted and measured TOC and $S_1 + S_2$ for the comparison of OM abundance of different source rocks in the Sikeshu Sag.

extending northward to the Chepaizi Uplift and northeast to the Shawan Sag. The Sikeshu Sag can be subdivided into northern slope, central depression, and southern piedmont areas.

2.2. Tectonic Evolution and Stratigraphy. The tectonic evolution processes of the Sikeshu Sag can be divided into three stages: (i) the initial subsidence stage (T_3-J_2): from the Late Hercynian to the Middle Yanshanian, strong tectonic sub-

https://doi.org/10.1021/acsomega.2c07991
ACS Omega 2023, 8, 10314−10334

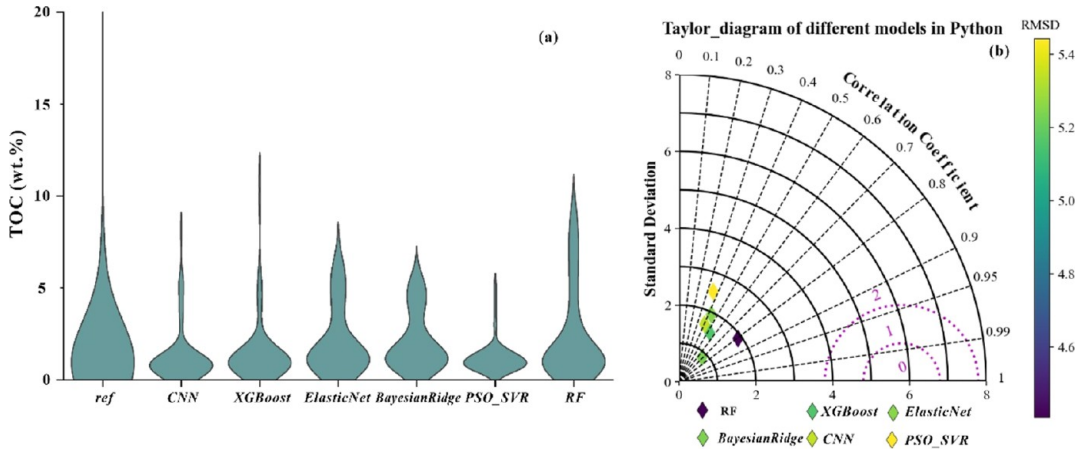


Figure 5. (**a**) Violin plot and (**b**) Taylor diagram for visual comparison of RF, PSO_SVR, XGBoost, ElasticNet, Bayesian Ridge, and CNN.

Table 1. Summary of Some Metrics Were Used to Evaluate the Performance of These Models Used in This Study

metrics	measured	RF	XGBoost	CNN	BR	ElasticNet	PSO_SVR
RMSD		0.61	0.61	0.68	0.71	0.74	0.78
R^2		0.81	0.45	0.42	0.60	0.44	0.36
Σ	6	1.9	1.5	1.7	2.0	0.9	2.6
max	20	11.5	12.5	9.0	8.0	8.5	6.0
performance		good	moderate	moderate	poor	poor	poor

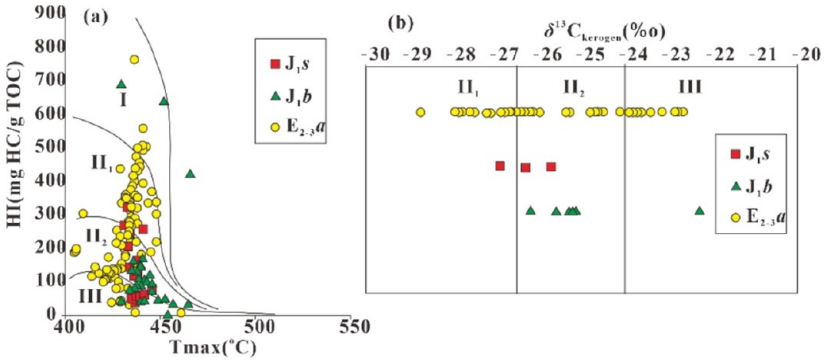


Figure 6. (**a**) Cross-plots of T_{max} versus HI and (**b**) scatter plot of the stable carbon isotopes of kerogen for the comparison of kerogen type in different source rocks from the Sikeshu Sag.

sidence and NS-trending compression led to the formation of the Kayindike anticline (Figure 1b,c). (ii) Alternate stages of aborted and continued slow stratum subsidence (J_2-E): from the Late Yanshanian to the Early Himalayan, the stratum subsidence was aborted; therefore, the Middle Jurassic Toutunhe Formation and the Upper Jurassic are missing from the Sikeshu Sag; from the Cretaceous to Paleogene, the Sikeshu Sag experienced long-term alternating stages of aborted and continued slow stratum subsidence; meanwhile, a regional unconformity surface was formed in the top area of the Mesozoic; (iii) rapid stratum subsidence stage (E-Q): influenced by the Himalayan Movement, slow and rapid subsidence alternately led to the formation of the Dushanzi and Xihu anticlines and the massive sets of Neogene Shawan, Taxihe, and Dushanzi formations in the Chepaizi Uplift.[31]

The stratigraphy of the Sikeshu Sag lacks the Permian deposits. The study area consists of four intervals: the Upper Jurassic Badaowan (J_1b), Sangonghe (J_1s), Middle Jurassic Xishanyao (J_2x), and Paleogene Anjihaihe ($E_{2\text{-}3}a$) formations. J_2x is mainly composed of tuffaceous mudstone, basalt, andesite, and rhyolite, with minor siliciclastic sedimentary rocks.[49−51] The Triassic to Neogene intervals are mainly interlayers or alternate deposits of alluvial fan-fluvial-lake facies, and the lithology is dominated by gray-black breccia and gray, gray-green, brown-red, and gray-brown conglomerate, siltstone, and mudstone (Figure 2).

3. MATERIALS AND METHODS

3.1. Sample Selection and Preparation. To investigate the differential hydrocarbon generation potential of $E_{2\text{-}3}a$ and J_1b source rocks and the oil−source discrimination of oil samples from different tectonic belts, several rock and oil samples were collected from the Anjihaihe outcrops and wells in different tectonic belts (Figure 1a). The following analyses were conducted: total organic carbon (TOC) content, Rock-Eval pyrolysis, vitrinite reflectance, column chromatography, and gas chromatography−mass spectrometry (GC−MS); previously published data were used,[52−54] and new data were collected from Shengli Oil Company, Sinopec.

https://doi.org/10.1021/acsomega.2c07991
ACS Omega 2023, 8, 10314−10334

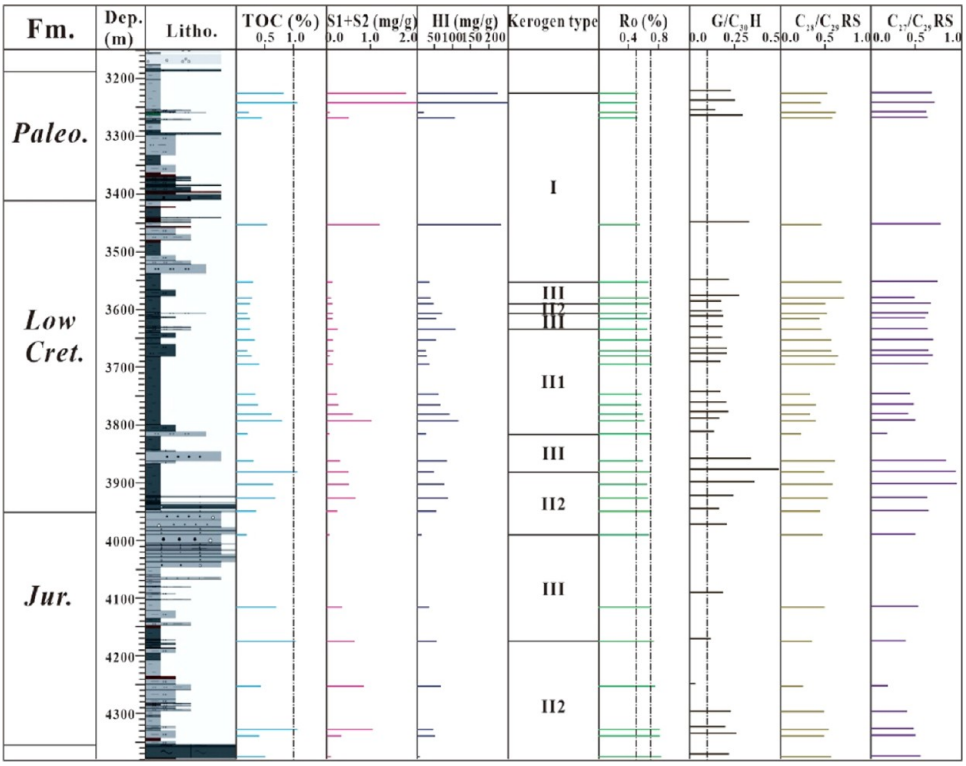


Figure 7. Comprehensive geochemical profile of Well Ka 6 in the Sikeshu Sag.

3.2. Gas Chromatography−Mass Spectrometry and GC−MS/MS Analyses. First, the source rock samples were cleaned, crushed, and powdered to 40−80 mesh size. OM was extracted using a LabTech high-speed solvent extraction instrument (HPSE Gemini EVS). First, the source rock samples were ground to less than 0.5 mm (40−80 mesh) and mixed with a certain amount of diatomite (bake at 200 °C for 4 h) until the ratio with the sample (dry, not suitable for baking) was approximately 5:1; then, the sample was placed in the extraction tank. The remaining space of the extraction tank was filled with diatomite to 0.5−1.0 cm from the upper edge to complete the installation of the extraction tank. The partial pressure gauge of the nitrogen valve was controlled at 0.8 MPa. Leach saturated hydrocarbons with *n*-hexane, 2:1 dichloromethane/*n*-hexane leached aromatic hydrocarbons, and eluted non-hydrocarbons with chloroform. Gas chromatography−mass spectrometry (GC−MS) analysis of the oil-bearing sandstone samples was performed using an Agilent 7890A GC/5977MSD instrument equipped with an HP-5MS fused silica capillary column (60 m × 0.25 mm × 0.25 μm). GC coupled with tandem MS (GC−MS/MS) measurements were conducted on an Agilent Intuvo 9000 GC system using a triple quadrupole mass detector TSQ 8000 Evo.

3.3. Organic Petrographic Analysis. To measure the vitrinite reflectance (Ro) and analyze the maceral composition, organic petrographic examination of 39 samples was conducted from five wells using a Zeiss Axioplan II microscope following the laboratory methods described by Taylor (2001). Maceral analysis was performed using a LABORLUX 12 POL fluorescence microscope. More than 300 points were counted for each sample.[55]

3.4. Quantitative Grain Fluorescence Analysis and Microthermometry of Fluid Inclusions. The sandstone core samples were smashed and washed with water to obtain typical grains as the analysis samples and then processed with dichloromethanerogen peroxide and hydrochloric acid in sequence. The samples were oven-dried and weighed using an electronic balance (JY manufactured in China). The fluorescence intensity was measured using a Varian Fluorospectro Photometer (Cary Eclipse, USA) with a fluorescence wavelength of 300−600 nm.[51−54,56]

Microthermometry of the primary fluid inclusions was performed using a calibrated Linkam THMS-600G Heating/freezing Stage. The total homogenization temperature (Th) of the primary fluid inclusions was obtained by cycling to the liquid phase. The measurements were performed at a heating rate of 10 °C/min.

3.5. Machine Learning Procedure. Previous studies have proposed many methods for calculating hydrocarbon generating potential parameters using ML methods. However, for any one specified ML approach, there coexisted advantages and disadvantages. These artificial intelligence methods need further optimization in evaluating the application effect. In this study, different models use the same well logging curves combination served as input parameters (LLS, LLD, DEN, GR, and SP). In this study, numerous ML algorithms, including RF, CNN, XGBoost, ElasticNetCV, Bayesian Ridge regression, and PSO-SVR, were executed by Anaconda3 (Python 3.9). The relevant code, if requested, will be provided.

4. RESULTS AND DISCUSSION

4.1. Hydrocarbon Generation Potential Evaluations by a Machine Learning Procedure. *4.1.1. Organic Matter Abundance Evaluation.* OM abundance indicators reflecting hydrocarbon source rocks such as TOC and chloroform bitumen "A" indicate three sets of potential source rocks developed in the Sikeshu Sag: the Lower Jurassic Badaowan (J_1b), Sangonghe (J_1s), and Paleogene Anjihaihe ($E_{2\text{-}3}a$)

https://doi.org/10.1021/acsomega.2c07991
ACS Omega 2023, 8, 10314−10334

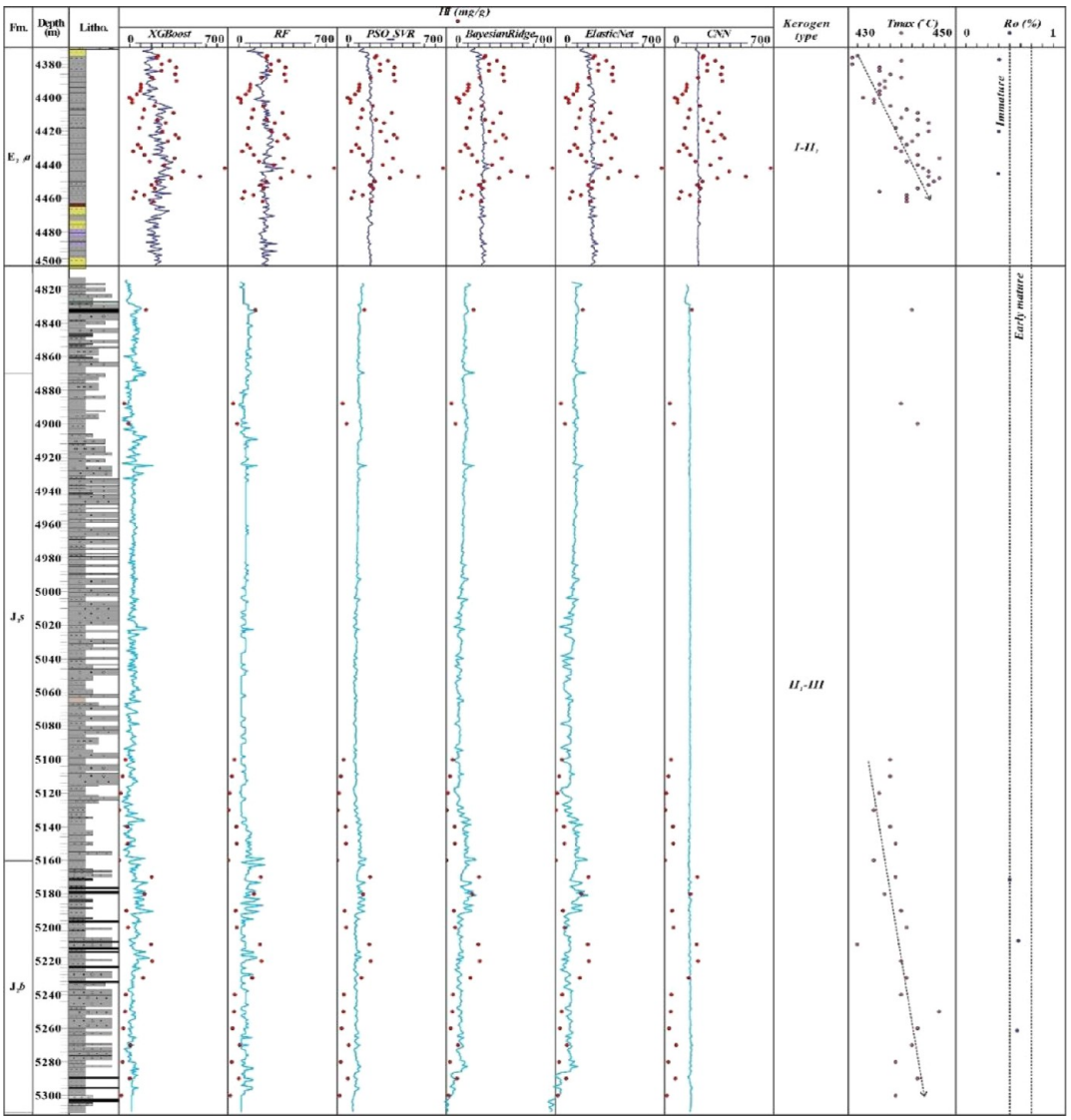


Figure 8. Comprehensive histogram of predicted and measured HI and measured T_{max} to compare the kerogen type and thermal maturity of different source rocks in the Sikeshu Sag.

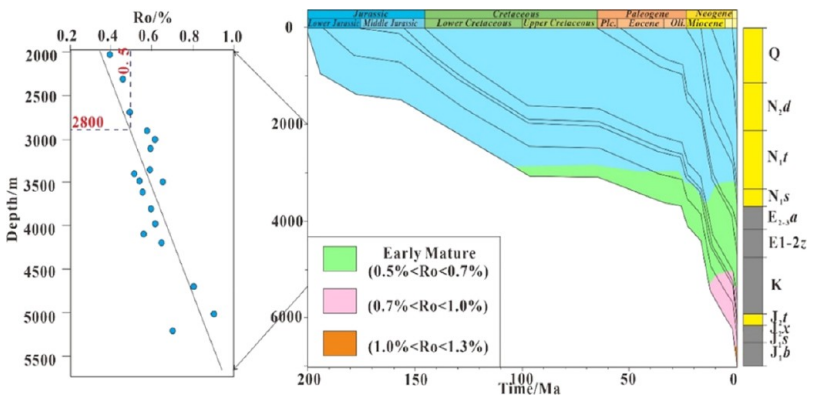


Figure 9. (**a**) Cross-plots of Ro versus burial depth and (**b**) burial-thermal history of Well Xihu-1 demonstrating the thermal evolution history of different source rocks in the Sikeshu Sag.

formations. A comprehensive investigation of the organic geochemistry shows that the J_1s source rock is a good source rock with moderate potential. The $E_{2\text{-}3}a$ and J_1b source rocks were of good quality. This study focuses on the difference in the hydrocarbon generation potential of the Jurassic and Paleogene source rocks.

The J_1b source rocks in the Sikeshu Sag are developed sets of black coal seams, gray-black carbonaceous mudstone, and dark

https://doi.org/10.1021/acsomega.2c07991
ACS Omega 2023, 8, 10314−10334

Table 2. Comprehensive Evaluation of OM Abundance, Type, and Maturity of Source Rocks in Different Strata

		OM abundance				kerogen types		thermal maturity		
Fm.		TOC (wt %)	chloroform bitumen A (%)	$S_1 + S_2$ (mg HC/g TOC)	classification	HI (mg HC/g rock)	classification	Ro (%)	T_{max} (°C)	classification
$E_{2\text{-}3}a$	min−max	0.4−1.42	0.012−0.3019	0.09−3.24	good	8.76−766.80	I−II_2	0.48−1.2	404.47−448.05	early mature
	average	0.97	0.07	1.39		254.4		0.64	431.63	
J_1b	min−max	0.37−10.64	0.007−0.32	0.07−5.3	good	0.88−170.89[a]	II_1−III	0.52−1.08	429.41−464.37	early mature−mature
	average	2.77	0.041	2.31		88.15		0.74	442.57	
J_1s	min−max	0.45−4.2	0.007−0.039	0.05−1.29	moderate-good	40.31−326.00	II_2−III	0.52−0.62	431.26−446.05	early mature
	average	0.99	0.026	0.76		141.22		0.53	436.96	

[a]Except for three outliers.

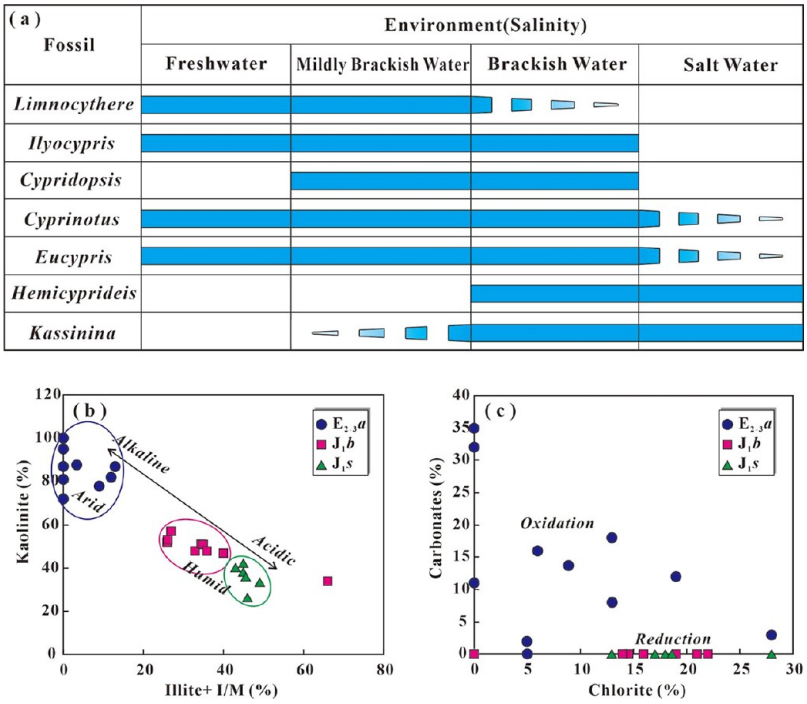


Figure 10. Ecological habits of the main ostracods in the $E_{2\text{-}3}a$ source rocks of the Sikeshu Sag.

mudstone. In terms of geochemical characteristics, the TOC content ranges from 0.37 to 10.64 wt %, with an average of 2.77 wt % (Figure 3a). The measured value of chloroform bitumen "A" is 0.007−0.32% (average on 0.041%, Figure 3b). The content of $S_1 + S_2$ was in the range of 0.07−5.3 mg/g, with an average of 2.31 mg/g.

The $E_{2\text{-}3}a$ source rock is composed of massive gray-green and gray mudstone intercalated with marl, limestone, and thin-layered sandstone, as well as shell limestone and a small amount of calcareous sandstone. In terms of geochemical characteristics, the TOC content ranges from 0.40 to 1.42 wt %, with an average of 0.97 wt %. The measured value of chloroform bitumen "A" is 0.012−0.3019% (average on 0.07%). The content of S_1+S_2 is located in the range of 0.09−3.24 mg/g, with an average of 1.39 mg/g. All these results indicate that these two sets of source rocks exhibit equivalent OM abundances. The J_1s source rock is relatively poor compared to the two sets of source rocks, with the TOC content ranging from 0.45 to 4.2 wt %, with an average of 0.99 wt %. The measured value of chloroform bitumen "A" is 0.007−0.039% (average on 0.026%). The content of $S_1 + S_2$ is placed in the range of 0.05−1.29 mg/g, with an average of 0.76 mg/g (Figure 3c,d).

To evaluate the abundance of OM in the vertical direction, numerous ML algorithms, including RF, CNN, SVR, XGBoost, ElasticNetCV, Bayesian Ridge regression, and PSO-SVR, were utilized to obtain a continuous TOC change curve at depth (Figure 4). A direct comparison of the model results is shown in Figure 4, where the continuous TOC and $S_1 + S_2$ values of $E_{2\text{-}3}a$ are lower than those of J_1s and J_1b, further indicating that the OM of $E_{2\text{-}3}a$ is worse than that of the Jurassic source rocks.

Taylor diagrams and violin plots were used to visually compare the performance of the models used in this study. A Taylor diagram is a polar plot that summarizes multiple aspects of model performance in a single diagram.[55] It uses three statistics to evaluate the degree of correspondence between the estimated and measured values: correlation coefficient (R^2), root-mean-squared error (RMSE), and standard deviation (σ). This is based on the following cosine relationship

$$\sigma^2 = r^2 + \mathrm{RMSE}^2 - 2 \times r \times \mathrm{RMSE} \times \cos\theta$$

https://doi.org/10.1021/acsomega.2c07991
ACS Omega 2023, 8, 10314−10334

Table 3. Summary of Mineral Contents Showing the Difference of $E_{2\text{-}3}a$ and J_1b Source Rocks

Fm.	lithology	tremolite (%)	zeolite (%)	quartz (%)	potash feldspar (%)	albite (%)	calcite (%)	dolomite (%)	clay minerals (%)	chlorite (%)	I/M (%)	illite (%)	kaolinite (%)
$E_{2\text{-}3}a$	gray-green mudstone	0	12	10	0	4	0	0	74	5	85	10	0
$E_{2\text{-}3}a$	purple-red mudstone	0	20	7	0	4	2	0	67	5	80	15	0
$E_{2\text{-}3}a$	gray-green mudstone	0	11	16	0	12	8	3	50	0	90	10	0
$E_{2\text{-}3}a$	gray-green mudstone	0	9	6	0	3	0	32	50	0	90	10	0
$E_{2\text{-}3}a$	red mudstone	0	0	26	0	9	6	2	57	13	63	15	9
$E_{2\text{-}3}a$	green mudstone	0	0	16	0	6	11	1	66	19	71	10	0
$E_{2\text{-}3}a$	red mudstone	0	0	22	0	6	13	3	56	6	72	10	12
$E_{2\text{-}3}a$	red mudstone	0	0	25	0	7	18	0	50	13	77	10	0
$E_{2\text{-}3}a$	green mudstone	0	0	23	0	9	3	0	65	28	62	10	0
$E_{2\text{-}3}a$	gray-yellow mudstone	0	0	17	0	7	35	0	41	0	77	10	13
$E_{2\text{-}3}a$	gray-yellow mudstone	0	5.2	16.8	0	6.7	9.6	4.1	57.6	8.9	76.7	11	3.4
J_1b	light green mudstone	0	0	36	0	0	0	0	64	0	29	5	66
J_1b	gray-green mudstone	0	0	40	0	8	0	0	52	22	47	5	26
J_1b	yellow-green siltstone	0	0	38	0	7	0	0	55	19	38	10	33
J_1b	gray-black mudstone	0	0	46	0	9	0	0	45	16	47	10	27
J_1b	gray-black mudstone	0	0	42	0	3	0	0	55	16	43	5	36
J_1b	gray-black siltstone	0	0	37	0	5	0	0	58	21	43	10	26
J_1b	gray-green mudstone	0	0	38	0	6	0	0	56	14	41	10	35
J_1b	gray-black mudstone	0	0	37	0	9	0	0	54	13	37	10	40
J_1b	gray-black mudstone	0	0	38.8	0	7.4	0	0	53.8	14.5	42.6	8.5	34.4
J_1s	dark brown mudstone	0	0	39	0	16	0	0	45	18	18	15	49
J_1s	green mudstone	0	0	28	0	23	0	0	49	13	32	10	45
J_1s	gray-green thin-bedded mudstone	0	0	27	0	12	0	0	61	17	23	15	45
J_1s	green mudstone	0	0	26	0	12	0	0	62	17	25	15	43
J_1s	gray mudstone	0	0	23	7	16	0	0	54	28	11	15	46
J_1s	gray mudstone	0	0	28.6	1.4	15.8	0	0	54.2	18.6	21.8	14	45.6

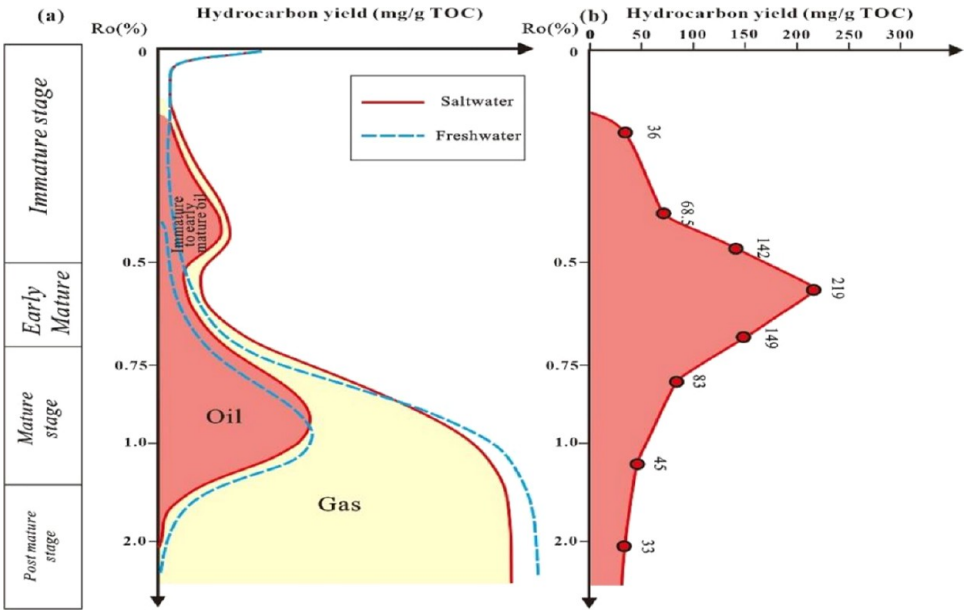


Figure 11. (**a**) Schematic diagram of the difference in hydrocarbon generation processes between saltwater and freshwater sediments and (**b**) hydrocarbon generation thermal simulation of the Anjihaihe outcrop in the Sikeshu Sag.

A violin plot is a boxplot with the addition of rotated kernel density on each side.[56] The violin plot (Figure 5a) clearly indicates the better accuracy of the RF and XGBoost models in contrast to the Bayesian Ridge, CNN, SVR, ElasticNetCV, and PSO-SVR models.

In addition, a Taylor diagram provides a statistical summary of how well the patterns match and is widely used to compare outputs from different models. Figure 5b shows a comparison of the predicted TOC values from different models and the measured TOC values. First, the correlation coefficients for the radial line denote the relationship between the measured TOC values and models, indicating similarities in the depth scales. Second, the normalized standard deviations (standard deviation of the model divided by that of the observations) on the *x*- and *y*-

https://doi.org/10.1021/acsomega.2c07991
ACS Omega 2023, 8, 10314−10334

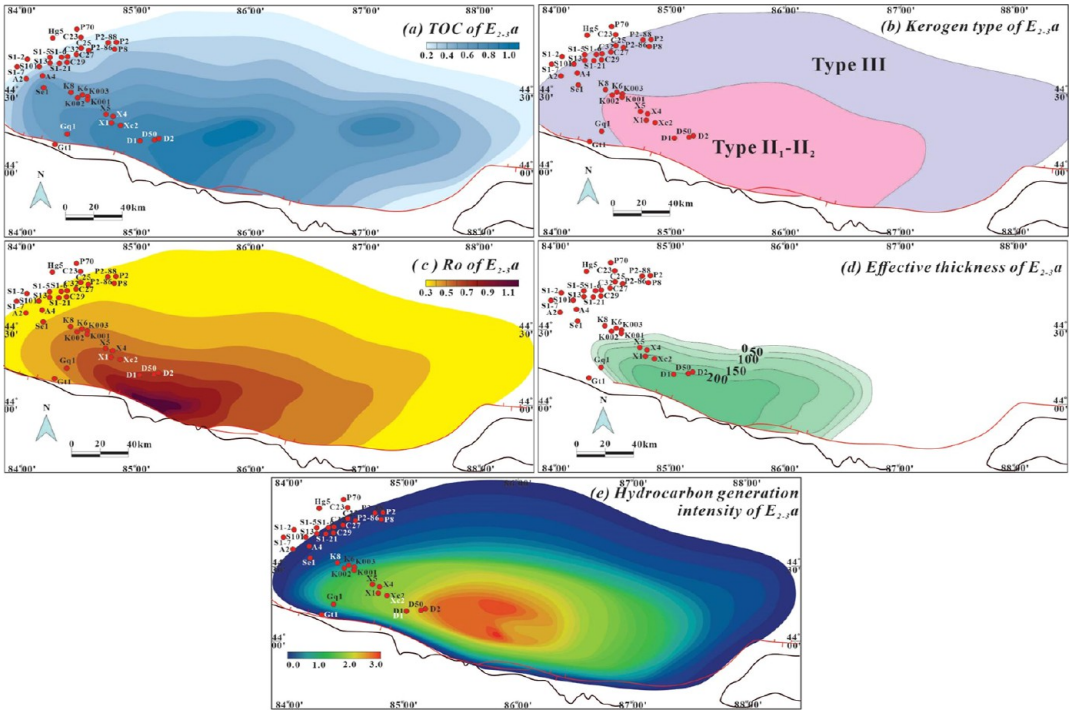


Figure 12. Contour of (a) TOC, (b) kerogen type, (c) Ro, (d) effective thickness, and (e) hydrocarbon generation intensity of $E_{2-3}a$ in the southwestern Junggar Basin.

axes indicate whether the variance of the model is larger or smaller than that in the measured TOC values. The points inside the dashed curved line indicate that the model simulations have lower variability than the measured TOC values. In contrast, the models with normalized standard deviations were higher than the measured TOC values. By combining the correlation and standard deviation, the points closer to the reference point on the *x*-axis show better agreement, making the ensemble mean of the RF a realistic model represent the true TOC value line. Comparing the performance of the ML model using the Taylor diagram (Figure 5b), the RF model yielded the highest performance (with low value of RMSE = 4.6%, high $R^2 > 0.91$, and $\sigma = 1.9$) of the six applied ML models (Table 1).

4.1.2. Kerogen Type. In this study, the cross-plots of T_{max} versus hydrogen index (HI) and the scatter plot of the stable carbon isotope of kerogen were utilized for the comparison of kerogen types of different source rocks in the Sikeshu Sag. As illustrated in Figure 6a,b, the kerogen types of J_1b are mainly type II_1, though several are type III. In contrast, $E_{2-3}a$ is mainly type I-II_2, and J_1s is mainly II_2-III. In addition, the comprehensive geochemical profile of Well Ka 6 in the Sikeshu Sag reveals that the source rocks of $E_{2-3}a$ exhibit relatively higher values of HI than those of J_1b and J_1s, and the ratios of C_{27}/C_{29} regular steranes (RS) for $E_{2-3}a$ exhibit higher values than those of J_1b and J_1s, implying that the kerogen type of $E_{2-3}a$ is better than that of J_1b and J_1s (Figure 7).

A direct comparison of the model results is shown in Figure 8, where the predicted and measured HI values of $E_{2-3}a$ are higher than those of J_1s and J_1b, further indicating that the kerogen type of $E_{2-3}a$ is better than that of the Jurassic source rocks.

4.1.3. Thermal Maturity. The comprehensive geochemical profile of Well Ka 6 in the Sikeshu Sag also reveals that the source rocks of $E_{2-3}a$ exhibit higher values of Ro than those of J_1b, implying that the kerogen type of $E_{2-3}a$ is better than that of J_1b and J_1s (Figure 8). The corresponding Ro is 0.52−1.08%, with an average of 0.74%, indicating that most OM from J_1b has entered the mature stage, and a certain amount has entered peak hydrocarbon generation, which is a mature and good source rock. Based on the statistical analysis of drilled holes and outcrops, the hydrocarbon generation threshold depth (Ro = 0.5%) of the J_1b source rock is approximately 2800 m (Figure 9), and this set of source rocks has already entered the stage of large-scale hydrocarbon generation. In contrast, the Ro values of the $E_{2-3}a$ and J_1s source rocks are still 0.5−0.7%, indicating that most of the OM from J_1b is still in the early mature stage (Table 2).

4.2. Early Mature Hydrocarbon Generation Controlled by Salt Lake Facies. The ostracod fossils in the $E_{2-3}a$ interval are mainly *Limnocythere dahangouensis*, *Ilyocypris cornae*, *Cyprinotus* (*Heterocypris*) *fortilis*, *Eucypris bella*, and *Austrocypris* (*Kassinina*) sp.; the Charophyte fossils mainly are *Maedleri-sphaera* and *Stephanochara*. From the analysis of paleontological habits in the outcrop samples, the $E_{2-3}a$ sedimentary period exhibited a brackish to saltwater sedimentary environment (Figure 10). The hydrocarbon-generating parent materials of source rocks in saline environments are mostly halophilic bacteria and algae. Water bodies in a saline environment have layered salinity and strong reducibility, providing favorable preservation conditions for OM and a high burial efficiency (23−25.9%). Moreover, OM is enriched in many layers of saline environments.[57] All of these serve as the material basis for efficient hydrocarbon generation in a saline environment. Furthermore, higher ratios of Sr/Ba (0.68−1.12, Figure 2) and $G/C_{30}H$ (Figure 7) and lower ratios of Pr/Ph imply that the environment of $E_{2-3}a$ features a water environment with hypersaline non-marine sediments.

The lithology of $E_{2-3}a$ is dominated by red and purplish-red mudstone, which are more oxidizing than the J_1b-J_1s sedimentary environment and in oxidizing conditions; the clay mineral assemblage is dominated by the I/M mixed layer and illite, reflecting the alkaline and arid sedimentary environment.

https://doi.org/10.1021/acsomega.2c07991
ACS Omega 2023, 8, 10314−10334

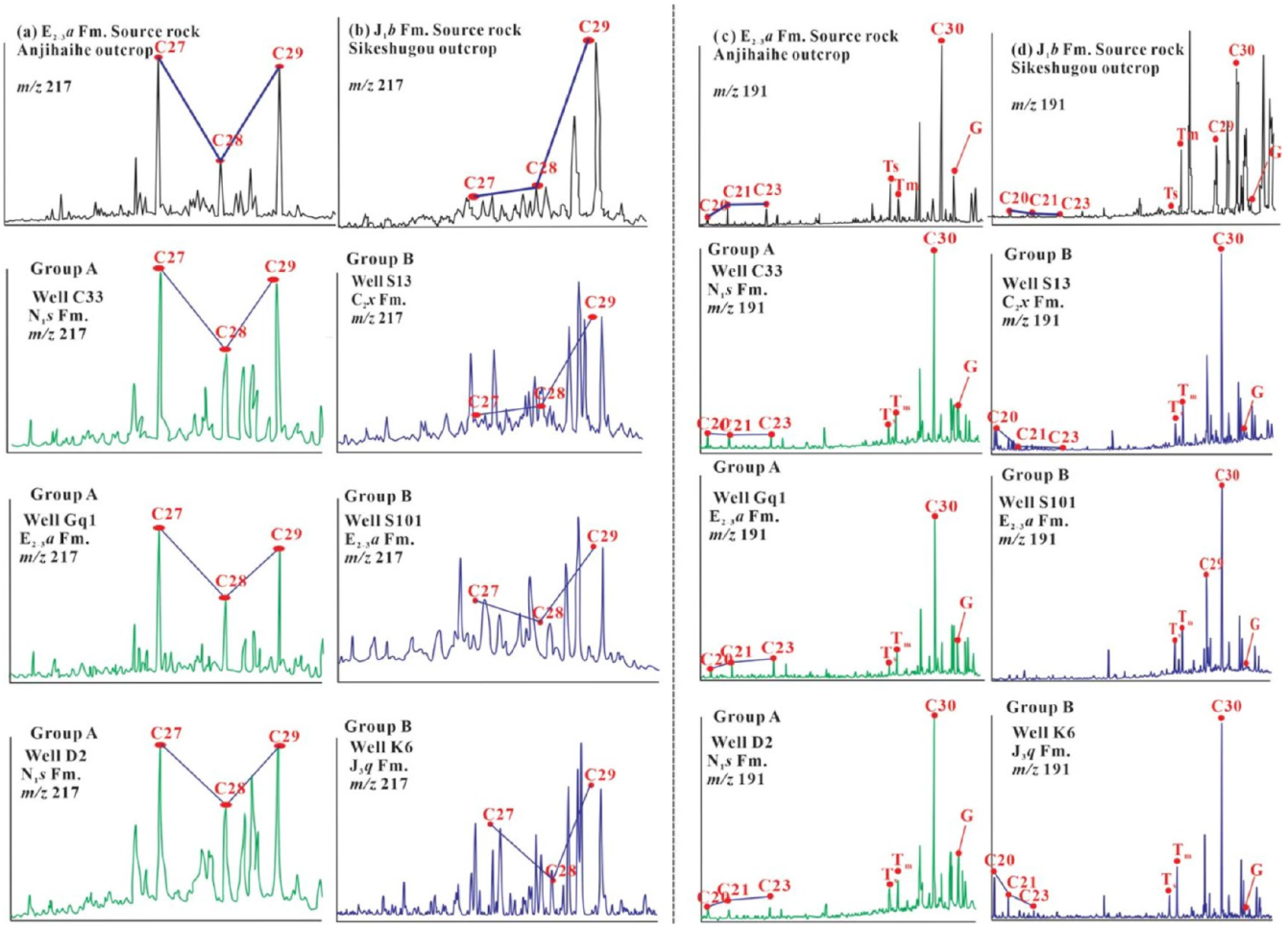


Figure 13. Representative mass chromatograms of m/z 217 and 191 showing the terpane and sterane characteristics of the J_1b and $E_{2\text{-}3}a$ source rocks and oils from different families (groups A and B) in the Sikeshu Sag.

The lithology of J_1b-J_1s is dominated by light-green and gray-black mudstones with high kaolinite content, reflecting the acidic conditions of the sedimentary water bodies and the warm and humid climate (Table 3).

Utilizing the coexistence model helps researchers avoid overestimating hydrocarbon generation and expulsion in a freshwater environment while underestimating hydrocarbon generation and expulsion in a saltwater environment. Extensive literature implies that the brackish water to saltwater transition with appropriate salinity (Cl^-) exhibits a distinct enhancement in the formation of low-maturity oil in the early thermal stage (Figure 11a).[32,36] In this study, thermal evolution simulation of the black mudstone from the Anjihaihe outcrop also showed that the high-quality source rocks formed in the saline lake basin can generate and expel hydrocarbons on a large scale in the early maturity stage (Figure 11b).

From the paleothermal evolution,[58] the similar geothermal fields of the Paleogene and Jurassic intervals imply that the corresponding hydrocarbon generation threshold depth of $E_{2\text{-}3}a$ should be similar to that of the Jurassic interval. Furthermore, coupled with an analysis of the burial history of the Sikeshu Sag, the oil generation threshold is approximately 3000−3500 m. However, drilling and seismic data inferred that the maximum burial depth of $E_{2\text{-}3}a$ exceeded 5000−6000 m, far greater than the hydrocarbon generation threshold. Source rocks also have the potential for large-scale hydrocarbon generation and expulsion. Its effective source rock area is approximately 1600 km^2, and it is primarily distributed in the deep depression area of the Sikeshu Sag.

Combining the aforementioned analysis of hydrocarbon generation history and the study of hydrocarbon generation kinetics, in this study, the source rock quality (TOC, Ro, thickness of black shale (H), and intensity of hydrocarbon generation (Q)) was comprehensively considered.

$$Q = H \times \rho \times TOC \times PR_{max} \times x$$

where **Q** is the hydrocarbon generation intensity (t/km^2), **H** is the thickness of the black shale (m), ρ is the density of the source rock (2.60 g/cm^3), and TOC is the total organic carbon content (wt %); $\mathbf{PR_{max}}$ is the maximum hydrocarbon production rate (mg/g TOC); and $\boldsymbol{x}$ is the conversion rate (%).

$E_{2\text{-}3}a$ is in the southeastern area of the Sikeshu Sag, whose source rocks are characterized by high OM abundance (TOC >1.0 wt %) and early-mature stage (0.5% < Ro< 0.7%); the thickness of black shale is more than 150 m, and the corresponding hydrocarbon generation intensity (Q) is higher than 2.0×10^4 t/km^2, which is the center of the $E_{2\text{-}3}a$ source kitchen located at the southeastern part of the Sikeshu Sag (Figure 12).

4.3. Oil Source Discrimination by Biomarkers and a Stable Carbon Isotope. Figure 13 shows the representative mass chromatograms (m/z 217 and 191) of steranes and

https://doi.org/10.1021/acsomega.2c07991
ACS Omega 2023, 8, 10314−10334

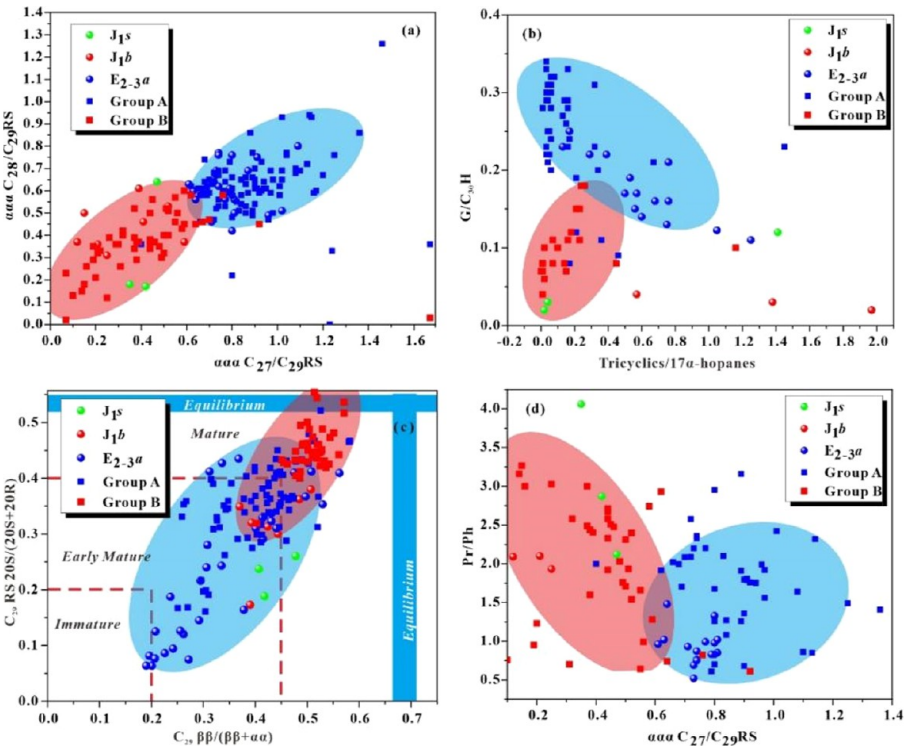


Figure 14. Cross-plot of (**a**) $\alpha\alpha\alpha C_{27}/C_{29}$ RS versus $\alpha\alpha\alpha C_{28}/C_{29}$ RS, (**b**) Tricyclics/17α-hopanes versus $G/C_{30}H$, (**c**) C_{29} $\beta\beta/(\beta\beta + \alpha\alpha)$ versus C_{29} 20S/(20S + 20R), and (**d**) $\alpha\alpha\alpha C_{27}/C_{29}$ RS versus Pr/Ph of crude oils and source rock extracts displaying the correlations of different source rocks and crude oils in the Sikeshu Sag.

Table 4. Summary of Geochemical Parameters Showing the Difference of $E_{2\text{-}3}a$ and J_1b Source Rocks

parameters	$E_{2\text{-}3}a$	J_1b-J_1s
$G/C_{30}H$	higher (0.11−0.31)	lower (0.02−0.22)
C_{20}, C_{21}, C_{23} TT distribution	ascending pattern	declining pattern
C_{27}, C_{28}, C_{29} steranes' distribution	"L" shape	"inverse L" shape
C_{19}/C_{21}TT	lower (average: 1.41)	higher (average: 4.24)
$\alpha\alpha\alpha C_{28}/C_{29}$ RS	higher (0.42−0.80)	lower (0.17−0.64)
$\alpha\alpha\alpha C_{27}/C_{29}$ RS	higher (0.61−1.09)	lower (0.12−0.59)
Pr/Ph	lower (0.52−1.48)	higher (1.94−4.06)
tricyclics/17α-hopanes	lower (0.13−1.25)	higher (0.02−1.97)
$\delta^{13}C$	heavier (−30.23‰ to 26.47‰)	lighter (−30.74 to 27.49‰)
$C_{29}\beta\beta/(\beta\beta + \alpha\alpha)$	lower (0.19−0.44)	higher (0.37−0.53)
C_{29}20S/(20S + 20R)	lower (0.06−0.33)	higher (0.17−0.45)

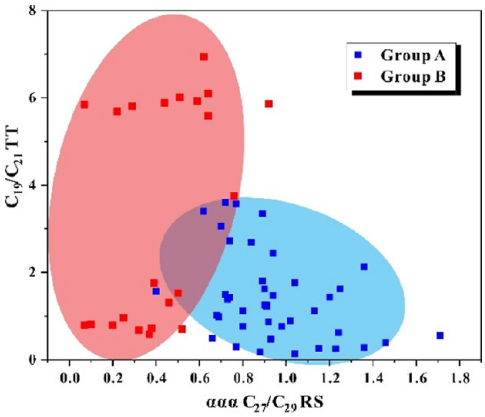


Figure 15. Cross-plot of $\alpha\alpha\alpha C_{27}/C_{29}$ RS versus C_{19}/C_{21} TT of crude oils displaying the correlations of different groups of crude oils in the Sikeshu Sag.

terpanes of multiple-source rock extracts and crude oils in the Sikeshu Sag. The C_{20}, C_{21}, and C_{23} tricyclic terpane (TT) distribution of $E_{2\text{-}3}a$ features an ascending pattern, and steranes C_{27}, C_{28}, and C_{29} exhibit an "L" shape. The C_{20}, C_{21}, and C_{23} TT distributions of the Jurassic (J_1b and J_1s) exhibit a declining pattern, and steranes C_{27}, C_{28}, and C_{29} exhibit an ascending shape.

Based on the comprehensive geochemical analysis and oil source discrimination results, the oils in different tectonic locations were classified into groups A and B, which were derived from the $E_{2\text{-}3}a$ and J_1b sources, respectively. Group A oils are mainly distributed in the Chunguang oilfield (Wells C2-200, C22, C27, C32, C33, and C27-1) and the Dushanzi oilfield (Wells D1, D2, Ds1, and D50), while group B oils are mainly distributed in the Kayindike anticline (Wells K001, K2, K003, and K6), the Xihu oil-bearing anticline (Wells X1 and Xc2), the Aika Fault Belt (Wells S13, S1-2, S3, and S103), and the Gaoquan anticline (Wells Gt1).

$\alpha\alpha\alpha$20R C_{27}RS/C_{29}RS and $\alpha\alpha\alpha$20R C_{28}RS/C_{29}RS are two source-related parameters (Figure 14a) that indicate a general increase in the relative content of C_{27}−C_{28} steranes and a decrease in C_{29} steranes in bacteria- and algae-enriched environments. The abundant C_{28} RS (higher ratios of $\alpha\alpha\alpha$20R C_{28}RS/C_{29}RS in the main range of 0.5−0.9) and C_{27} RS (higher ratios of $\alpha\alpha\alpha$20R C_{27}RS/C_{29}RS in the main range of 0.6−1.4) in the source rock extracts of $E_{2\text{-}3}a$ from the Sikeshu Sag and crude oils of group A suggest clay-rich source rocks deposited phytoplankton input.[59,60] Gammacerane is abundant in hypersaline non-marine sediments and related oils.[61] All samples of

https://doi.org/10.1021/acsomega.2c07991
ACS Omega 2023, 8, 10314−10334

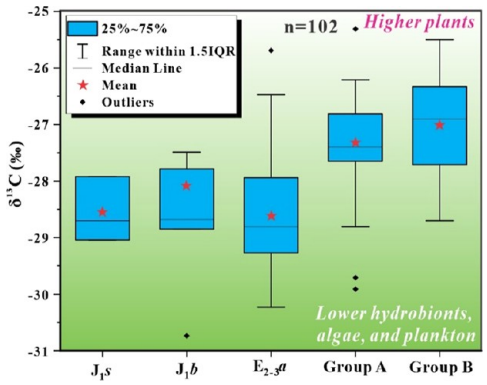


Figure 16. Box chart of $\delta^{13}C$ showing the correlations of different source rocks and crude oils in the Sikeshu Sag.

$E_{2\text{-}3}a$ from the Sikeshu Sag and crude oils of group A contain a high gammacerane index ($G/C_{30}H$) in the range of 0.15−0.35, which also suggests that the crude oils of group A were derived from the saline−brackish sediments in $E_{2\text{-}3}a$ (Figure 14a,b and Table 4). The opposite phenomenon appeared in the Jurassic (J_1b and J_1s) of the Sikeshu Sag, and crude oils of group B samples showed a lower content of C_{27} RS compared with C_{29} homologues (mainly distributed in the range of 0−0.6) and lower content of C_{28} RS compared with C_{29} homologues (mainly distributed in the range of 0.5−1.0), both indicating a higher input of terrigenous freshwater OM. Thus, the crude oils of group B were mainly derived from the Jurassic strata (J_1b, J_1s) of the Sikeshu Sag.

Tricyclics/17α-hopanes are primarily source-related parameters that compare a group of bacterial or algal lipids (tricyclics) with markers that arise from different prokaryotic species (hopanes), tricyclics/17α-hopanes also increase with increasing thermal maturity.[62] As shown in Figure 14c, the cross-plots of tricyclics/17α-hopanes versus $G/C_{30}H$ also confirm that the crude oils of group A are derived from the $E_{2\text{-}3}a$ of the Sikeshu Sag, and the crude oils of group B were derived from the Jurassic strata (J_1b, J_1s) of the Sikeshu Sag.

As shown in Figure 14c, the ratios of C_{29} $\beta\beta/(\beta\beta + \alpha\alpha)$ and C_{29} 20S/(20S + 20R) of $E_{2\text{-}3}a$ source rocks are lower than those of J_1b source rocks, and the corresponding ratios of group A and B oils are located in the zones of $E_{2\text{-}3}a$ and Jurassic (J_1b, J_1s) source rocks, which also confirm their genetic affinities.

Compared to the J_1b source rocks, the $E_{2\text{-}3}a$ source rocks feature lower ratios of short-chain C_{19} TT to their long-chain C_{21} counterparts (C_{19}/C_{21}TT) and lower ratios of Pr/Ph (Figures 14d and 15).

Tissot and Welte (1984) proposed that kerogens derived from higher plants are heavy,[34] with $\delta^{13}C$ ranging from −28.0 to −23.0‰, while the $\delta^{13}C$ of kerogens originating from lower hydrobionts, algae, and plankton ranges from −32.0 to −28.0‰. For $\delta^{13}C$, oils that differ by more than 2−3‰ are typically from different sources,[63] and oils of similar maturity differ by no more than 1‰. In the study area, the distribution of $\delta^{13}C$ values of the oils of group A (−29.90‰ to −25.30‰, average of −27.34‰) are similar to the $\delta^{13}C$ value of the $E_{2\text{-}3}a$ source rocks in the Sikeshu Sag (−30.23 to −26.47‰, average of −28.72‰), which revealed that the crude oils of group A were genetically related to the lower hydrobionts, algae, and plankton that dominated $E_{2\text{-}3}a$ in the Sikeshu Sag. The $\delta^{13}C$ values of the crude oils of group B range from −28.70 to −25.50‰ (average: 27.00‰, Figure 16, Table 5), which is consistent with the $\delta^{13}C$ value of the Jurassic (J_1b, J_1s) source rocks in the Sikeshu Sag (Figure 16, −30.74 to −27.49‰, average of −28.70‰).

Table 5. Summary of Stable Carbon Isotope Values of Crude Oil and Source Rock Extracts

oil/source	$\delta^{13}C$	oil/source	$\delta^{13}C$	oil/source	$\delta^{13}C$	oil/source	$\delta^{13}C$
J_1s	−28.70	$E_{2\text{-}3}a$	−28.78	D1	−28.80	P8-c3	−26.60
J_1s	−27.92	$E_{2\text{-}3}a$	−28.92	D1	−25.30	C2-p41	−26.70
J_1s	−29.04	$E_{2\text{-}3}a$	−29.24	D1	−26.30	C27	−27.30
J_1b	−28.85	$E_{2\text{-}3}a$	−29.35	D2	−26.20	C32	−27.80
J_1b	−28.85	$E_{2\text{-}3}a$	−29.84	D2	−27.80	K001	−25.50
J_1b	−27.49	$E_{2\text{-}3}a$	−30.23	D2	−27.10	K001	−26.40
J_1b	−30.74	$E_{2\text{-}3}a$	−29.11	D2	−27.20	K002	−25.90
J_1b	−28.60	$E_{2\text{-}3}a$	−29.04	D2	−27.14	K6	−27.72
J_1b	−28.08	$E_{2\text{-}3}a$	−29.40	D50	−27.82	K6	−28.70
J_1b	−28.76	$E_{2\text{-}3}a$	−28.70	D50	−27.60	K6	−26.20
J_1b	−23.38	$E_{2\text{-}3}a$	−28.96	X4	−27.48	K6	−26.80
$E_{2\text{-}3}a$	−25.69	$E_{2\text{-}3}a$	−28.00	X4	−26.50	K003	−26.26
$E_{2\text{-}3}a$	−26.47	$E_{2\text{-}3}a$	−28.52	X4	−26.30	X5	−26.00
$E_{2\text{-}3}a$	−27.17	$E_{2\text{-}3}a$	−27.94	Xc2	−26.40	S13	−26.20
$E_{2\text{-}3}a$	−27.91	$E_{2\text{-}3}a$	−27.71	Xc2	−26.30	S3	−26.60
$E_{2\text{-}3}a$	−28.20	$E_{2\text{-}3}a$	−28.57	Gq1	−27.60	S3	−27.70
$E_{2\text{-}3}a$	−28.81	D1	−27.64	Gq1	−27.39	S3	−26.64
$E_{2\text{-}3}a$	−29.87	D1	−26.84	Gq1	−27.36	P2-93	−27.40
$E_{2\text{-}3}a$	−29.95	D1	−27.53	C27	−27.40	C2-p1	−28.70
$E_{2\text{-}3}a$	−29.53	D1	−27.65	P2-18	−26.80	C2-p2	−28.32
$E_{2\text{-}3}a$	−29.24	D1	−27.88	P2-40	−26.80	P70	−27.40
$E_{2\text{-}3}a$	−29.27	D1	−27.62	P2-86	−27.80	P70	−27.80
$E_{2\text{-}3}a$	−29.11	D1	−27.56	P2-p87	−28.10	C26	−27.00
$E_{2\text{-}3}a$	−28.47	D1	−27.55	P2-88	−27.60	Gt1	−26.80
$E_{2\text{-}3}a$	−27.92	D1	−29.90	P206	−26.80		
$E_{2\text{-}3}a$	−27.51	D1	−29.70	P8-p1	−27.10		

https://doi.org/10.1021/acsomega.2c07991
ACS Omega 2023, 8, 10314−10334

Table 6. Selected Biomarkers of Oils and Source Rock Extracts in Different Strata in the Sikeshu Sag for Oil Source Discrimination

samples	C_{27}/C_{29}RS	C_{28}/C_{29}RS	TT/H	Pr/Ph	G/C_{30}H	$C_{29}\beta\beta/(\beta\beta + \alpha\alpha)$	C_{29}20S/(20S + 20R)	original group	predicted group	P(D > d |G = g)	discriminant function
J_1s	0.35	0.18	0.02	4.06	0.02	0.42	0.19	B	B	0.999	−2.663
J_1s	0.42	0.17	0.04	2.87	0.03	0.41	0.24	B	B	0.996	−2.248
J_1s	0.47	0.64	1.41	2.12	0.12	0.48	0.26	B	B	0.915	−1.242
J_1b	0.12	0.37	1.97	2.09	0.02	0.39	0.17	B	B	1.000	−3.686
J_1b	0.21	0.36	0.57	2.10	0.04	0.42	0.31	B	B	1.000	−3.139
J_1b	0.25act	0.31	1.38	1.94	0.03	0.39	0.32	B	B	1.000	−3.007
J_1b	0.15	0.50			0.04	0.37	0.35	B	B	1.000	−3.403
J_1b	0.41	0.46			0.12	0.48	0.36	B	B	0.971	−1.593
J_1b	0.59	0.37			0.18	0.51	0.38	B	A[a]	0.712	−0.230
J_1b	0.39	0.61			0.22	0.44	0.30	B	B	0.741	−0.833
J_1b	0.52	0.53			0.07	0.53	0.45	B	B	0.958	−1.469
$E_{2\text{-}3}a$	0.61	0.63	1.24	0.96	0.11	0.20	0.06	A	B[a]	0.627	−0.668
$E_{2\text{-}3}a$	0.63	0.60	1.25	1.02	0.11	0.19	0.06	A	B[a]	0.544	−0.563
$E_{2\text{-}3}a$	0.64	0.56	0.56	1.48	0.15	0.20	0.08	A	A	0.737	−0.191
$E_{2\text{-}3}a$	0.71	0.46	0.76	0.93	0.16	0.21	0.08	A	A	0.895	0.151
$E_{2\text{-}3}a$	0.80	0.42	0.75	0.98	0.13	0.23	0.09	A	A	0.930	0.290
$E_{2\text{-}3}a$	0.81	0.50	0.68	1.02	0.16	0.27	0.07	A	A	0.975	0.622
$E_{2\text{-}3}a$	0.77	0.56	0.60	0.99	0.14	0.24	0.09	A	A	0.930	0.289
$E_{2\text{-}3}a$	0.81	0.56	0.57	0.85	0.17	0.21	0.13	A	A	0.978	0.658
$E_{2\text{-}3}a$	0.80	0.58	0.57	0.83	0.17	0.26	0.12	A	A	0.979	0.679
$E_{2\text{-}3}a$	0.79	0.61	0.53	0.83	0.19	0.26	0.13	A	A	0.983	0.734
$E_{2\text{-}3}a$	0.74	0.62	0.50	0.87	0.17	0.29	0.14	A	A	0.950	0.401
$E_{2\text{-}3}a$	0.73	0.67	0.29	0.52	0.22	0.24	0.19	A	A	0.984	0.764
$E_{2\text{-}3}a$	0.73	0.69	0.13	0.69	0.23	0.29	0.22	A	A	0.987	0.819
$E_{2\text{-}3}a$	0.74	0.77	0.39	0.76	0.22	0.38	0.16	A	A	0.985	0.790
$E_{2\text{-}3}a$	0.74	0.76	0.76	0.75	0.21	0.34	0.24	A	A	0.983	0.735
$E_{2\text{-}3}a$	0.80	0.76	0.17	1.33	0.25	0.31	0.24	A	A	0.997	1.303
$E_{2\text{-}3}a$	0.91	0.75			0.26	0.31	0.28	A	A	1.000	1.881
$E_{2\text{-}3}a$	1.09	0.80			0.28	0.41	0.29	A	A	1.000	2.896
$E_{2\text{-}3}a$	1.02	0.51			0.29	0.40	0.31	A	A	1.000	2.638
$E_{2\text{-}3}a$	0.96	0.49			0.29	0.44	0.31	A	A	1.000	2.353
$E_{2\text{-}3}a$	0.87	0.69			0.25	0.43	0.32	A	A	0.999	1.624
$E_{2\text{-}3}a$	0.72	0.64			0.31	0.41	0.33	A	A	0.998	1.425
D1	0.80	0.48	0.32	1.26	0.23	0.25	0.17	A	A	0.995	1.112
D1	0.79	0.73	0.16	0.61	0.29	0.27	0.17	A	A	0.999	1.562
D1	0.90	0.70	0.13	1.36	0.27	0.30	0.16	A	A	1.000	1.916
D1	0.97	0.69	0.15	1.56	0.26	0.31	0.19	A	A	1.000	2.148
D1	1.08	0.70	0.17	1.64	0.24	0.30	0.20	A	A	1.000	2.559
D1	1.10	0.69	0.32	0.86	0.31	0.30	0.22	A	A	1.000	3.140
D1	1.01	0.58		2.42	0.29	0.34	0.26	A	A	1.000	2.536
D1	0.97	0.59		1.92	0.28	0.37	0.29	A	A	1.000	2.354
D2	0.40	0.36	0.16	2.00	0.28	0.45	0.37	A	A	0.621	−0.356
D2	0.90	0.65	0.67	0.68	0.21	0.44	0.40	A	A	0.998	1.404
D2	0.84	0.54	0.10	1.08	0.33	0.45	0.36	A	A	0.997	1.403
Ds1	1.02	0.69	0.04		0.29	0.45	0.41	A	A	1.000	2.589
Ds1	0.93	0.73	0.21		0.19	0.40	0.41	A	A	0.998	1.472
Ds1	0.98	0.65	0.04		0.31	0.40	0.42	A	A	1.000	2.576
D50	1.16	0.59			0.29	0.48	0.43	A	A	1.000	3.309
Gq1	0.96	0.47		1.99	0.21	0.50	0.43	A	A	0.999	1.765
Gq1	0.89	0.51		3.16	0.18	0.50	0.46	A	A	0.995	1.126
X4	1.13	0.72	0.16	0.85	0.33	0.50	0.48	A	A	1.000	3.486
Xc2	0.94	0.64	0.05		0.31	0.51	0.47	A	A	1.000	2.412
Xc2	1.15	0.93	0.06		0.20	0.58	0.47	A	A	1.000	2.473
Xc2	1.36	0.86	0.14		0.29	0.47	0.36	A	A	1.000	4.241
Xc2	1.04	0.76	0.16		0.23	0.43	0.39	A	A	1.000	2.213
Xc2	0.77	0.69	0.04		0.25	0.47	0.37	A	A	0.995	1.111
Xc2	1.20	0.67	0.03		0.29	0.46	0.36	A	A	1.000	3.462
Xc2	0.88	0.86	0.36		0.11	0.41	0.32	A	A	0.973	0.597
Gq1	1.15			1.90	0.28	0.44	0.40	A	A	1.000	3.486
Gq1	0.96	0.47		1.99	0.21	0.45	0.41	A	A	1.000	2.433
Gq1	0.89	0.51		3.16	0.18	0.40	0.41	A	A	1.000	2.483

https://doi.org/10.1021/acsomega.2c07991
ACS Omega 2023, 8, 10314−10334

Table 6. continued

samples	C_{27}/C_{29}RS	C_{28}/C_{29}RS	TT/H	Pr/Ph	G/C_{30}H	$C_{29}\beta\beta/(\beta\beta+\alpha\alpha)$	C_{29}20S/(20S + 20R)	original group	predicted group	P(D > d |G = g)	discriminant function
C2-200	1.23	0.00	0.97		0.68	0.31	0.20	A	A	1.000	6.693
P2-86	0.92	0.59	0.08	1.76	0.32	0.52	0.37	A	A	1.000	2.411
C22	0.80	0.61	0.06	1.68	0.32	0.42	0.36	A	A	1.000	1.849
C27	0.91	0.61	0.06	1.79	0.31	0.45	0.35	A	A	1.000	2.282
C32	0.91	0.61	0.06	1.81	0.31	0.58	0.47	A	A	1.000	2.226
C33	0.90	0.61	0.06	1.80	0.30	0.45	0.35	A	A	1.000	2.187
P19-1	1.24	0.33	0.21		0.12	0.34	0.35	A	A	1.000	2.281
P8	0.94	0.61	0.01	1.75	0.28	0.33	0.33	A	A	1.000	2.205
P28	0.66	0.46	0.34	2.02	0.20	0.33	0.31	A	A	0.919	0.238
C2-5	1.71	0.36	0.46		0.09	0.33	0.32	A	A	1.000	4.261
C16-1	1.46	1.26	1.45		0.23	0.32	0.31	A	A	1.000	4.228
P2-18	0.89	0.77	0.03	1.91	0.34	0.49	0.43	A	A	1.000	2.381
P2-40	0.77	0.64	0.03	2.20	0.33	0.42	0.37	A	A	0.999	1.784
P2-86	0.68	0.74	0.06	2.00	0.28	0.46	0.43	A	A	0.992	0.972
P2-p87	0.69	0.62	0.06	1.70	0.31	0.46	0.43	A	A	0.997	1.251
P2-88	0.80	0.22	0.17	2.95	0.08	0.30	0.68	A	A	0.823	−0.036
P206	0.62	0.62	0.03	1.92	0.22	0.44	0.45	A	A	0.913	0.216
P8-p1	0.70	0.65	0.03	2.09	0.23	0.44	0.41	A	A	0.976	0.627
P8-c3	0.72	0.64	0.03	2.09	0.30	0.38	0.40	A	A	0.998	1.339
C2-p41	0.74	0.67	0.04	2.36	0.21	0.35	0.39	A	A	0.980	0.689
C27	0.73	0.68	0.05	2.20	0.22	0.53	0.52	A	A	0.984	0.770
C32	0.72	0.68	0.05	2.58	0.25	0.44	0.41	A	A	0.990	0.910
C27-1	0.74	0.69	0.06	2.33	0.24	0.44	0.43	A	A	0.991	0.941
K001	0.92	0.45	0.45	0.61	0.08	0.43	0.42	B	A[a]	0.961	0.482
K001	0.76	0.58	0.44	0.82	0.11	0.49	0.40	B	B	0.999	−3.549
K6	0.51	0.52		1.93	0.08	0.45	0.41	B	B	0.947	−1.395
K6	0.38	0.29	0.25	1.60	0.30	0.54	0.47	B	B	1.000	−3.135
K6	0.44	0.50	0.25	2.58	0.09	0.55	0.48	B	B	1.000	−3.141
K003	0.55	0.57	1.16	0.64	0.10	0.34	0.34	B	B	0.869	−1.093
S13	0.07	0.23	0.00	4.48	0.07	0.35	0.36	B	B	1.000	−3.549
S1-2	0.29	0.36	0.02		0.06	0.35	0.35	B	B	0.999	−2.603
S1-2	0.14	0.15	0.01	3.16	0.08	0.40	0.32	B	B	1.000	−3.126
S3	0.15	0.18	0.01	3.27	0.07	0.49	0.46	B	B	1.000	−3.151
S3	0.16	0.26	0.01	3.00	0.04	0.50	0.46	B	B	1.000	−3.354
S3	0.22	0.33	0.01		0.08	0.50	0.50	B	B	0.999	−2.788
C50	0.44	0.38	0.02	2.33	0.10	0.49	0.49	B	B	0.972	−1.597
P400-1	0.46	0.37	0.10	2.48	0.10	0.52	0.46	B	B	0.953	−1.435
P400-3	0.50	0.40	0.07	2.31	0.11	0.51	0.47	B	B	0.907	−1.210
P2-300	0.52	0.40	0.24	2.40	0.18	0.50	0.43	B	B	0.569	−0.594
C8	0.38	0.37	0.26	2.44	0.18	0.49	0.41	B	B	0.915	−1.241
C110E	0.37	0.39	0.23	2.49	0.15	0.48	0.40	B	B	0.962	−1.506
C114E	0.32	0.40	0.21	2.58	0.15	0.45	0.43	B	B	0.983	−1.755
P2-93	0.07	0.02	0.16	2.04	0.11	0.46	0.43	B	B	1.000	−3.227
C2-p1	0.20	0.21	0.23	1.23	0.11	0.48	0.43	B	B	0.999	−2.599
C2-p2	0.10	0.13	0.18	0.76	0.12	0.48	0.45	B	B	1.000	−2.973
P70	0.25	0.12	0.15	3.03	0.07	0.53	0.49	B	B	0.999	−2.656
P70	1.71	0.03	0.14		0.08	0.51	0.55	B	A[a]	1.000	4.221
C26	0.39	0.37	0.07	2.40	0.08	0.52	0.54	B	B	0.992	−2.004
C29	0.45			2.54	0.11			B	B	0.999	−2.588
Gt1	0.38			2.40	0.12			B	B	1.000	−2.967

[a]Represents wrongly predicted cases; TT/H: tricyclics/17α-hopanes).

4.4. Oil Source Discrimination by Stepwise Discriminant Analysis. Stepwise discriminant analysis (SDA) uses samples pre-classified into known groups (or classes) as learning samples to establish functions that best discriminate among classes. The resulting functions are then implemented to classify unknown samples. In oil−source correlation, known or learning samples represent those whose sources are pre-identified, and investigating samples represent those whose sources need to be identified.[64,65] For a more comprehensive identification of multiple oil sources of crude oils in the Sikeshu Sag, the data of the candidate geochemical parameters were tested for the normality of distribution using the Kolmogorov−Smirnov (K−S) test at a confidence level (α) of 5%. Most of the parameters, where $\alpha > 0.05$, were demonstrated to be log−normal distribution variables for each source rock class. Prior to SDA, the geochemical parameter candidate data were normalized for 87 source rocks and 52 oil samples using the average and standard deviation values of the source rock samples. The Wilks'

https://doi.org/10.1021/acsomega.2c07991
ACS Omega 2023, 8, 10314−10334

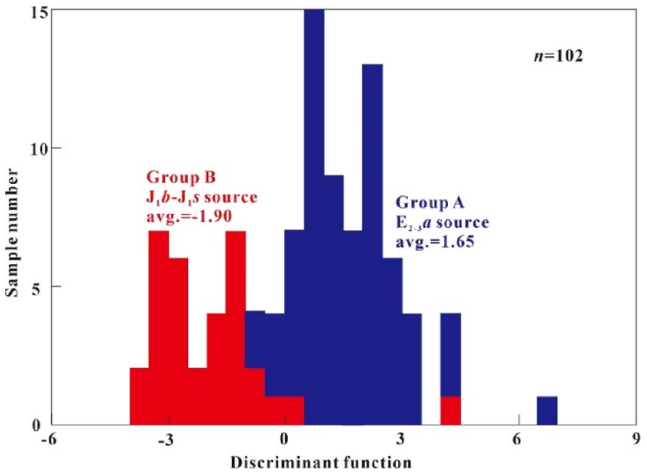


Figure 17. Histogram of the discriminant function of the oils from different families (groups A and B) derived from different oil sources in the Sikeshu Sag.

lambda significance level decreased to 0.014, which was significantly less than 0.05, and the correct rate of the original validation reached 95.5% (Table 6).

The results of SDA (Figure 17 and Table 6) indicate that the crude oils of group A were derived from $E_{2\text{-}3}a$ of the Sikeshu Sag, whereas those of group B were derived mainly from the Jurassic (J_1b, J_1s) of the Sikeshu Sag.

4.5. Determination of Oil Charging History Combined with Fluid Inclusion Microthermometry and One-Dimensional Basin Modeling. The homogenization temperatures (Th) of coeval aqueous inclusions in Well K6 (group B, J_1b source) ranged from 40.0 to 100.0 °C (Figure 18a), with a low main peak at 70−90 °C, while those in Well X5 (group A, $E_{2\text{-}3}$a source) ranged from 100.0 to 150.0 °C (Figure 18b), with a high main peak at 110−130 °C. Combined with the one-dimensional (1-D) basin modeling, the oil charging periods of group A and B oils are the Middle-Late Miocene (13−8 Ma) and Late Oligocene (23−20 Ma).

4.6. Hydrocarbon Migration Pathways by Quantitative Grain Fluorescence Analysis and Its Implications for Petroleum Exploration. Quantitative grain fluorescence (QGF) can be utilized to identify hydrocarbon migration pathways, that is, the fluorescent intensity from the adsorbed and captured hydrocarbons in the inclusion. The hydrocarbon migration pathways can be determined using the 0−4 QGF indices.[66−71]

As observed in Figure 19, the values of the QGF index and QGF-E in the upper section are higher than those of the lower section of Well D1, demonstrating that—$J_3q^3 - J_3q^2$ is a paleo-reservoir.[72,73]

In addition, the planar distribution of the QGF index shows that from the southeast region (Well Ds1) to the northwest region (Wells S13 and S103), the corresponding QGF index gradually increases, indicating that the hydrocarbon-bearing paleo-fluid flows from the southeast to the northwest (Figure 20 and Table 7)..

5. CONCLUSIONS

(1) The Sikeshu Sag exists as two sets of main source rocks: the Paleogene Anjihaihe ($E_{2\text{-}3}a$) and J_1b formations. The thermal maturity of J_1b has entered the "oil generation window", while the $E_{2\text{-}3}a$ source rocks are still in the early maturity stage; owing to its sedimentary background in a saline−brackish water environment, large-scale hydrocarbon generation and expulsion can be realized in the early maturity stage.

(2) The $E_{2\text{-}3}a$ source rock exhibits the characteristic "L" shape of C_{27}, C_{28}, and C_{29} sterane distribution, ascending pattern of the C_{20}, C_{21}, and C_{23} TT distribution, and higher ratios of gammacerane index, $\alpha\alpha\alpha C_{27}/C_{29}$ RS, and $\alpha\alpha\alpha C_{28}/C_{29}$ RS, and low ratios of C_{19}/C_{21}TT. The Jurassic strata (J_1b, J_1s) feature a declining pattern of C_{20}, C_{21}, and C_{23} TT distribution, an "inverse L" shape of C_{27}, C_{28}, and C_{29} steranes, a low gammacerane index, $\alpha\alpha\alpha C_{27}/C_{29}$ RS, $\alpha\alpha\alpha C_{28}/C_{29}$ RS, and high ratios of C_{19}/C_{21}TT.

(3) The comprehensive analysis of fluid inclusion microthermometry and 1-D basin modeling revealed that the oil

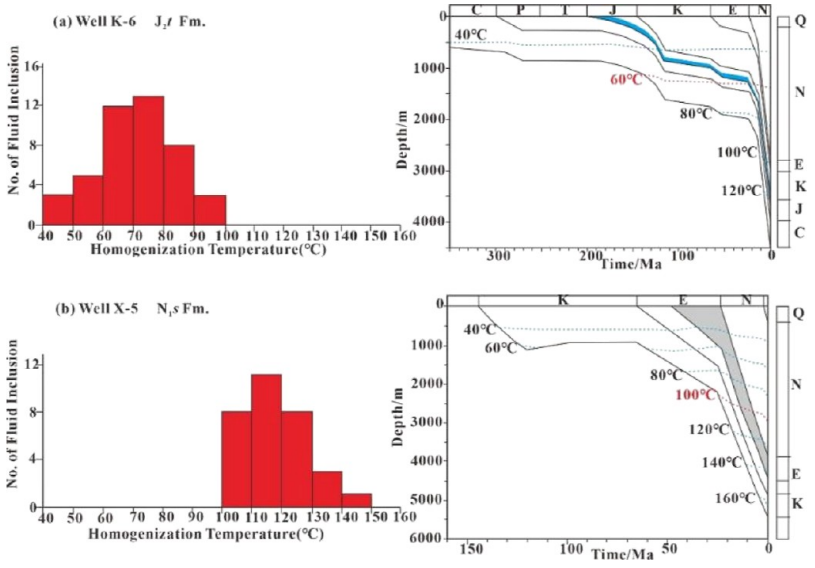


Figure 18. Histograms of homogenization temperature (Th) for coexisting aqueous inclusions of oil inclusions and thermal history of (a) J_2t formation in Well K6 and (b) N_1s formation in Well X5 in the Sikeshu Sag.

https://doi.org/10.1021/acsomega.2c07991
ACS Omega 2023, 8, 10314−10334

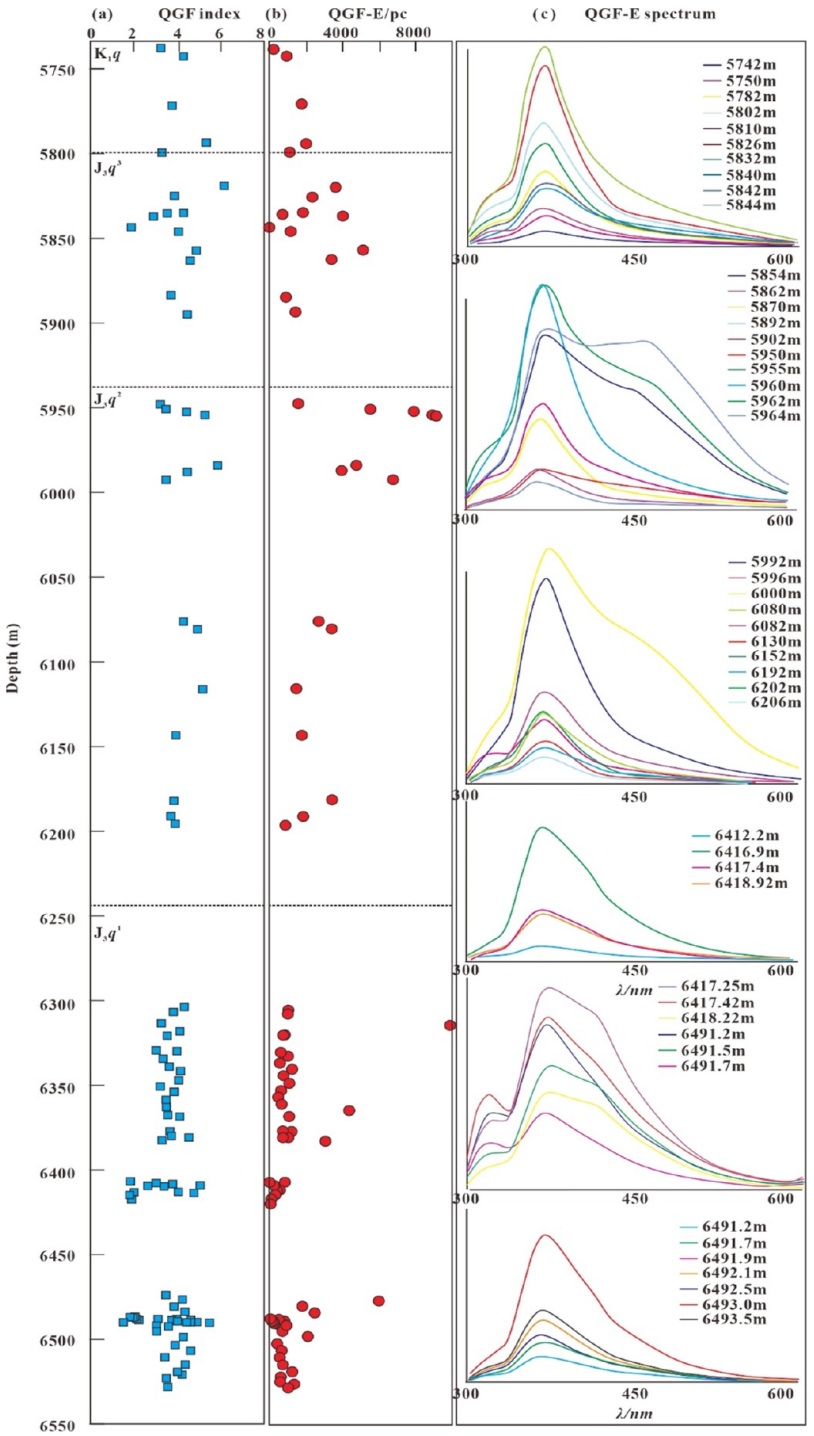


Figure 19. Vertical variations of the (**a**) QGF index, (**b**) QGF-E, and (**c**) QGF spectrum of Well D1 in the Sikeshu Sag demonstrating the vertical hydrocarbon migration direction.

charging periods of the group A and B oils are in the Middle-Late Miocene (13−8 Ma) and the late Oligocene (23−20 Ma), respectively.

(4) This study proposes a new understanding that N_1s oils in Western Chepaizi are derived from $E_{2\text{-}3}a$ low-mature saline lacustrine sediments and clarifies that the hydrocarbon generation potential and supply range of the $E_{2\text{-}3}a$ high-quality source rocks are located east of the Western Chepaizi Uplift and the interior area of the Sikeshu Sag. The oil−source correlation results also show that the oil reservoirs in the Sikeshu Sag and its peripheral region developed two sets of distinct source rocks, particularly the $E_{2\text{-}3}a$ source rocks, both of which have large-scale hydrocarbon generation and expulsion.

(5) To improve the optimization ability of the parameters, several optimization algorithms, such as adaptive particle swarm optimization, multi-population genetic algorithm, gray wolf optimization, ant colony algorithm, simulated annealing algorithm, and whale optimization algorithm, can be used to determine the optimal value of the parameters of the discriminant model and to realize the

https://doi.org/10.1021/acsomega.2c07991
ACS Omega 2023, 8, 10314−10334

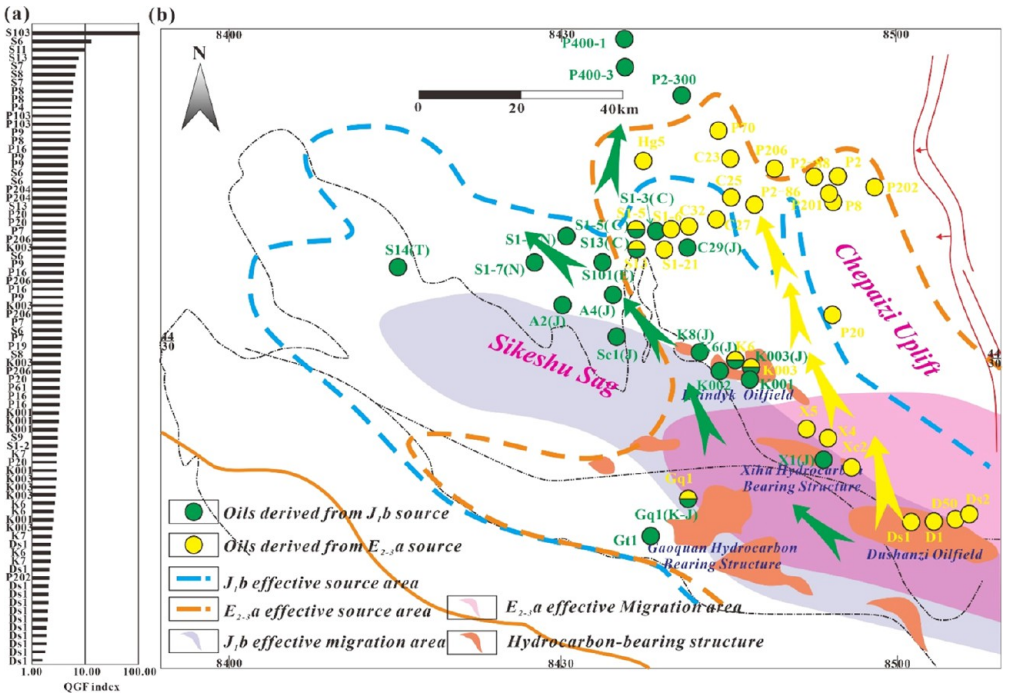


Figure 20. (**a**) Histogram of the QGF index and (**b**) schematic diagram showing the charging range of two different oil sources ($E_{2\text{-}3}a$ and J_1b-J_1s) of the southwestern Junggar Basin.

Table 7. Summary Table of the QGF Index for Different Well Samples

well	QGF index	well	QGF index	well	QGF index	well	QGF index
Ds1	1.54	K001	2.89	P7	3.68	S6	4.67
Ds1	1.80	P20	2.9	P206	3.69	P9	4.77
Ds1	1.86	K7	2.93	K003	3.77	P9	4.78
Ds1	1.87	S1-2	3.01	P9	3.83	P16	4.82
Ds1	1.89	S9	3.02	P16	3.84	P8	5.08
Ds1	1.90	K001	3.08	P206	3.85	P9	5.13
Ds1	1.92	K001	3.1	P16	3.87	P103	5.25
Ds1	2.02	K001	3.16	P9	3.88	P103	5.32
Ds1	2.03	P16	3.24	S6	4.15	P4	5.46
Ds1	2.07	P16	3.29	K003	4.23	P8	5.65
P202	2.11	P61	3.3	P206	4.3	P8	5.86
Ds1	2.16	P20	3.35	P7	4.35	S7	5.92
K7	2.19	P206	3.49	P20	4.38	S8	6.35
K6	2.25	K003	3.52	P20	4.38	S7	6.71
Ds1	2.26	S8	3.53	S13	4.40	S13	7.53
K7	2.35	P19	3.6	P204	4.45	S11	10.1
K003	2.42	P7	3.6	P204	4.52	S6	13.33
K001	2.53	S6	3.63	S6	4.58	S103	105.83
K6	2.55	K003	2.84	K003	2.8	K003	2.73
K6	2.65						

automatic optimization of model parameters, which can help enhance the prediction ability of the ML model.

AUTHOR INFORMATION

Corresponding Author

Tengfei Yu − *School of Geosciences, China University of Petroleum (East China), Qingdao 266580, China; Exploration and Development Research Institute, Shengli Oilfield Company, SINOPEC, Dongying 257000, China;* orcid.org/0000-0001-9768-2885; Email: yutengfei.slyt@sinopec.com

Authors

Zhiping Wu − *School of Geosciences, China University of Petroleum (East China), Qingdao 266580, China*

Ruichao Guo − *Exploration and Development Research Institute, Shengli Oilfield Company, SINOPEC, Dongying 257000, China*

Guanlong Zhang − *Exploration and Development Research Institute, Shengli Oilfield Company, SINOPEC, Dongying 257000, China*

Yuejing Zhang − *Exploration and Development Research Institute, Shengli Oilfield Company, SINOPEC, Dongying 257000, China*

Fengkai Shang − *Exploration and Development Research Institute, Shengli Oilfield Company, SINOPEC, Dongying 257000, China*

Lin Chen − *Exploration and Development Research Institute, Shengli Oilfield Company, SINOPEC, Dongying 257000, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.2c07991

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This study was funded by the Sinopec science and technology project (P20013). We thank the Shengli Oil Company of Sinopec for approving the publication. We also acknowledge the experts of Elsevier Webshop support for its linguistic assistance during the preparation of this manuscript.

NOMENCLATURE

RF	random forest
CNN	convolutional neural networks
XGBoost	extreme gradient boosting
ElasticNet BRR	Bayesian Ridge Regression
PSO-SVR	particle swarm optimization-support vector regression
TOC	total organic carbon
$S_1 + S_2$	hydrocarbon potential
HI	hydrogen index

REFERENCES

(1) Aydin, G.; Karakurt, I.; Aydiner, K. Analysis and Mitigation Opportunities of Methane Emissions from the Energy Sector. *Energy Sources, Part A* **2012**, *34*, 967−982.

(2) Aydin, G.; Karakurt, I.; Aydiner, K. Modeling of coal consumption in Turkey: An application of trend analysis. In *24th International Mining Congress and Exhibition of Turkey*: Antalya, Turkey, 2015; pp 83−87.

(3) Kocurek, K.; Martindale, R. C.; Day, D.; Goudge, T. A.; Kerans, K.; Hassenruck-Gudipati, H. J.; Mason, M.; Cardenas, B. T.; Petersen, E. I.; Mohrig, M.; Aylward, D. S.; Hughes, C. M.; Nazworth, C. M. Antecedent aeolian dune topographic control on carbonate and evaporite facies: Middle Jurassic Todilto Member, Wanakah Formation, Ghost Ranch, New Mexico, USA. *Sedimentology* **2019**, *66*, 808−837.

(4) Ji, H.; Sumei, L.; Paul, G.; Zhang, H. G.; Pang, X. Q.; Xu, T. W.; He, N. N.; Shi, Q. Geochemical characteristics and significance of heteroatom compounds in lacustrine oils of the dongpu depression (bohai bay basin, China) by negative-ion fourier transform ion cyclotron resonance mass spectrometry. *Mar. Pet. Geol.* **2018**, *97*, 568−591.

(5) Keighley, D. Phosphatic Carbonate Shale of the "Bird's Nest Saline Zone", Upper Green River Formation, Uinta Basin, Utah. *Stratigraphy and Paleolimnology of the Green River Formation*; Springer Netherlands: Western USA, 2015, pp 251−276.

(6) Qiu, Y.; Yang, Y. X.; Wang, X. L.; Wan, Y.; Hu, W. X.; Lu, J. J.; Tao, G. L.; Li, Z.; Meng, F. W. In situ Raman spectroscopic quantification of aqueous sulfate: experimental calibration and application to natural fluid inclusions. *Chem. Geol.* **2019**, *533*, 119447.

(7) Tänavsuu-milkeviciene, K.; Frederick Sarg, J. F. Evolution of an organic-rich lake basin - stratigraphy, climate and tectonics: Piceance Creek basin, Eocene Green River Formation. *Sedimentology* **2012**, *59*, 1735−1768.

(8) Tomasso, M.; Wright, W. R.; Costa, F. O.; Alfredo, D. A.; Marilia, V. S.; Elizabeth, M.; Michael, R. H.; Martin, P. J. Diversity in salt-tectonic structures and their relationships to evaporite facies and pre-evaporite depositional and structural regimes in the Campos Basin, offshore Brazil; *AAPG Bull.*, **2016**.

(9) Wu, Y. P.; Liu, C. L.; Jiang, F. J.; Hu, T.; Lv, J. H.; Zhang, C. X.; Guo, X. G.; Huang, L. L.; Hu, M. L.; Huang, R. D.; Awan, R. S.; Zhao, Y. Geological characteristics and shale oil potential of alkaline lacustrine source rock in Fengcheng Formation of the Mahu Sag, Junggar Basin, Western China. *J. Pet. Sci. Eng.* **2022**, *216*, 110823.

(10) Chen, J. P.; Wang, X. L.; Deng, C. P.; Zhao, Z.; Ni, Y. Y.; Sun, Y. G.; Yang, H. B.; Wang, H. T.; Liang, D. G.; Zhu, R. K.; Peng, X. L. Geochemical feature of source rocks in the southern margin, Junggar Basin, Northwestern China. *Acta Pet. Sin.* **2015**, *36*, 767−780 (in Chinese with English abstract).

(11) Chen, Y. F. Hydrocarbon Accumulation Law Research in Chepaizi Area of Junggar Basin. Master Degree Thesis, China University of Petroleum, 2011. (in Chinese with English abstract).

(12) Qin, J.; Lin, X. Y.; Pan, H.; He, B. Correlation Between the Sikeshu Sag and the Oil Source Rock in Its Peripheral Region, South Junggar Basin. *Bull. Mineral., Petrol. Geochem.* **2014**, *33*, 395−400. (in Chinese with English Abstract)

(13) Xiao, F.; Liu, L. F.; Zeng, L. Y.; Wu, K. J.; Xu, Z. J.; Zhou, C. X. Geochemical characteristics and oil source of crude oils in the east edge of Chepaizi high, Junggar Basin. *J. China Univ. Min. Technol.* **2014**, *43*, 646−655. (in Chinese with English Abstract)

(14) Cheng, C. L. Re-Cognition of the Jurassic Source Rocks in Sikeshu Sag. *Xinjiang Pet. Geol.* **2018**, *39*, 119−124. (in Chinese with English Abstract)

(15) Wei, L. Evaluation of Upper Triassic Hydrocarbon Source Rocks in Sikeshu Sag, the Southern Margin of Junggar Basin. *J. Xi'an Shiyou Univ.* **2018**, *33*, 9−15.

(16) Ma, W. Y.; Rouzi, D. D.; Li, J.; Liao, J. D.; Wang, H. J.; He, D. Hydrocarbon generation of Jurassic source rocks in the Southern Margin of Junggar Basin. *Xinjiang Pet. Geol.* **2020**, *41*, 31−37. (in Chinese with English abstract)

(17) Shi, X. P.; Qi, X. F.; Shi, Q. D.; Hu, Z. F.; Hou, L. Characteristics of Jurassic source rocks and oil source of Sikeshu Depression in the Junggar Basin. *Chin. J. Geol.* **2020**, *55*, 974−988. (in Chinese with English abstract)

(18) Zhang, Y. B.; Wang, M. J.; Du, Z. L.; Yuan, Q. D. Main oil source rocks and oil-source correlation in the Sikeshu Sag, the Southern Margin of Junggar Basin. *Mar. Geol. Front.* **2020**, *36*, 39−48. (in Chinese with English abstract)

(19) Du, J.; Zhi, D. M.; Li, J. Z.; Yang, D. S.; Tang, Y.; Qi, X. F.; Xiao, L. X.; Wei, L. Y. Major breakthrough of Well Gaotan 1 and exploration prospects of lower assemblage in southern margin of Junggar Basin, NW China. *Pet. Explor. Dev.* **2019**, *46*, 216−227.

(20) Haiqing, H.; Zhi, D. M.; Lei, D. W.; Yang, D. S.; Xiao, L. X.; Yuan, B.; Qi, X. F.; Zhao, J. Y. Strategic breakthrough in Gaoquan anticline and exploration assessment on lower assemblage in the southern margin of Junggar Basin. *China Petroleum Exploration* **2019**, *24*, 137−146.

(21) Zhang, Z. Y.; Zhu, G. Y.; Chi, L. X.; Wang, P. J.; Zhou, L.; Li, J. F.; Wu, Z. H. Discovery of the high-yield well GT1 in the deep strata of the southern margin of the Junggar Basin, China: Implications for liquid petroleum potential in deep assemblage. *J. Pet. Sci. Eng.* **2020**, *191*, 107178.

(22) Shi, B. B.; Chang, X. C.; Liu, Z. Q.; Pang, Y. M.; Xu, Y. D.; Mao, L. X.; Zhang, P. F.; Chen, G. Intelligent identification of sequence stratigraphy constrained by multipopulation genetic algorithm and dynamic time warping technique: A case study of Lower Cretaceous Qingshuihe Formation in hinterland of Junggar Basin (NW China). *Basin Research* **2021**, *33*, 2517−2544.

(23) Guo, C. Q. *Study on Oil-generating Potential of Sikeshu Sag, Junggar Basin*; Chengdu University of Technology Master Degree Thesis, 2005. (in Chinese with English abstract).

(24) Huang, Y. Q.; Qian, M. H.; Zhang, M. W. Comprehensive Analysis of Hydrocarbon Accumulation in the Southern Margin of Sikeshu Sag in Junggar Basin. *J. Oil Gas Technol.* **2009**, *31*, 7−11. (in Chinese with English Abstract)

(25) Liu, L. F.; Meng, J. H.; Wang, W. B.; Jin, J.; Wu, L.; Zhao, Y. D.; Wang, P.; Zhi, D. M. Differences in Geochemical Characteristics of Oils trapped in the Upper and the Lower series of strata of Chepaizi Uplift along Northwest Margin of Junggar Basin and their significances. *J. Jilin Univ.* **2011**, *41*, 377−390. (in Chinese with English abstract)

(26) Zhang, Z. H.; Liu, H. J.; Li, W.; Fei, J. J.; Xiang, K.; Qin, L. M.; Xi, W. J.; Zhu, L. Origin and Accumulation Process of Heavy Oil in Chepaizi Area of Junggar Basin. *J. Earth Sci. Environ.* **2014**, *36*, 18−32. (in Chinese with English Abstract)

(27) Zhang, Z. H.; Xiang, K.; Qing, L. M.; Zhuang, W. S.; Xi, W. J.; Zhao, S. F. Geochemical characteristics od source rocks and their contribution to petroleum accumulation of Chepaizi area in Sikeshu Depression, Junggar Basin. *Geol. China* **2012**, *39*, 326−337. (in Chinese with English abstract)

(28) Jin, J.; Wang, F. Y.; Ren, J. L.; Feng, W. P.; Ma, W. Y.; Li, S. J. Genesis of High-Yield Oil and Gas in Well Gaotan-1 and Characteristics of Source Rocks in Sikeshu Sag, Junggar Basin. *Xinjiang Pet. Geol.* **2019**, *40*, 145−151. (in Chinese with English Abstract)

(29) Ren, J. L.; Wang, F. Y.; Zhao, Z. Y.; Zhang, Y. P.; Mi, J. L.; Liu, C. M. Genesis of Oil and Gas in Sikeshu Sag in the Southern Margin of Junggar Basin. *Xinjiang Pet. Geol.* **2020**, *41*, 25−30. (in Chinese with English Abstract)

https://doi.org/10.1021/acsomega.2c07991
ACS Omega 2023, 8, 10314−10334

(30) Gao, J. Geochemical characteristics and potential evaluation of Cretaceous-Paleogene source rocks in the southern margin of Junggar Basin. *Geolog. Rev.* **2021**, *67*, 219−221. (in Chinese with English Abstract)
(31) Shi, B. B.; Chang, X. C.; Xu, Y. D.; Wang, Y. R.; Mao, L. X.; Wang, Y. Origin and migration pathway of biodegraded oils pooled in multiple-reservoirs of the Chepaizi Uplift, Junggar Basin, NW China: insights from geochemical characterization and chemometrics methods. *Mar. Pet. Geol.* **2020**, *122*, 104655.
(32) Shi, B. B.; Chang, X. C.; Xu, Y. D.; Mao, L. X.; Zhang, J. H.; Li, Y. Charging history and fluid evolution for the Carboniferous volcanic reservoirs in the western Chepaizi Uplift of Junggar Basin as determined by fluid inclusions and basin modelling. *Geological Journal* **2020**, *55*, 2591−2614.
(33) Zhu, M.; Liang, Z. L.; Ma, J.; Pang, Z. C.; Wang, J.; Jiao, Y. Patterns of hydrocarbon generation and reservoir distribution in the Jurassic strata, Sikeshu Sag, Junggar Basin. *Nat. Gas Geosci.* **2020**, *31*, 488−497. (in Chinese with English abstract)
(34) Tissot, B.; Welte, D. H. Geochemical Fossils and Their Significance in Petroleum Formation. *Petroleum Formation and Occurrence*; Springer: Berlin, 1984, pp 93−130.
(35) Guo, P.; Liu, C. Y.; Gibert, L.; Huang, L.; Zhang, D. W.; Dai, J. How to find high-quality petroleum source rocks in saline lacustrine basins: A case study from the Cenozoic Qaidam Basin, NW China. *Mar. Pet. Geol.* **2020**, *111*, 603−623.
(36) Xiao, Z. L.; Chen, S. J.; Li, Y.; Su, K. M.; He, Q. B.; Han, M. M. Local high-salinity source rock and origin of crude oil in the xianshuiquan structure in the northwestern Qaidam Basin, China. *J. Pet. Sci. Eng.* **2021**, *198*, 108233.
(37) Erdenetsogt, B. O.; Hong, S. K.; Choi, J. Y.; Lee, I. Depositional environment and petroleum source rock potential of Mesozoic lacustrine sedimentary rocks in central Mongolia. *Mar. Pet. Geol.* **2022**, *140*, 105646.
(38) Liu, S. J.; Gao, G.; Jin, J.; Zhang, L. L.; Gang, W. Z.; Xiang, B. L.; Wang, J.; Wang, M. Newstudy and significance of hydrocarbon generation characteristics of source rock samples dominated by telalginite and lamalginite in saline lacustrine sediments. *Acta Geol. Sin.* **2022**, DOI: 10.19762/j.cnki.dizhixuebao.2022015.
(39) Liu, S. J.; Gao, G.; Jin, J.; Gang, L. L.; Xiang, W. Z.; Xiang, B. L. Source rock with high abundance of C28 regular sterane in typical brackish-saline lacustrine sediments: Biogenic source, depositional environment and hydrocarbon generation potential in Junggar Basin, China. *J. Pet. Sci. Eng.* **2022**, *208*, 109670.
(40) Ma, L.; Song, M.; Wang, Y. S.; Wang, Y. G.; Liu, H. M. Exploration progress of the Paleogene in Jiyang Depression, Bohai Bay Basin. *Energy Geosci.* **2023**, *4*, 42−50.
(41) Wang, Z.; Luo, X.; Zhang, L. Q.; Lu, X. Y.; Lu, Y. Q. The Anjihaihe source rock characteristics and hydrocarbon accumulation condition in the southern edge of Junggar Basin. *China J. Geol.* **2010**, *45*, 256−267. (in Chinese with English Abstract)
(42) Yang, F. *Evaluation of Ziniquanzi reservoir in Sikeshu Sag in Junggar Basin*; Yangtze University, Master Degree Thesis, 2014. (in Chinese with English abstract).
(43) Yu, K. H.; Cao, Y. C.; Qiu, L. W.; Sun, P. P. The hydrocarbon generation potential and migration in an alkaline evaporite basin: The Early Permian Fengcheng Formation in the Junggar Basin, northwestern China. *Mar. Pet. Geol.* **2018**, *98*, 12−32.
(44) Yu, K. H.; Cao, Y. C.; Qiu, L. W.; Sun, P. P.; Jia, X. Y.; Wan, M. Geochemical characteristics and origin of sodium carbonates in a closed alkaline basin: The Lower Permian Fengcheng Formation in the Mahu Sag, northwestern Junggar Basin, China. *Palaeogeography, Palaeoclimatology, Palaeoecology* **2018**, *511*. DOI: 10.1016/j.palaeo.2018.09.015.
(45) Yu, K. H.; Cao, Y. C.; Qiu, L. W. Resource Potentials of Soda and Boron in the Lower Permian Fengcheng Formation of the Mahu Sag in Northwestern Junggar Basin, China. *Acta Geol. Sin.* **2019**, *93*(). DOI: 10.1111/1755-6724.13830.
(46) Zhang, X. J.; Zhang, Y. J.; Mu, Y. Q.; Shang, F. K.; Yang, D. G.; Chen, L. Analysis on sedimentary and evolution of Paleogene in Sikeshu Sag, Junggar Basin. *Complex Hydrocarbon Reservoirs* **2019**, *12*, 17−22. (in Chinese with English abstract)
(47) Zeng, L. F. Petroleum Generation Kinetics of Jurassic Coaly Source Rocks in Junggar Basin. Doctor Degree Thesis, University of Chinese Academy of Sciences, 2020. (in Chinese with English abstract).
(48) Liu, H. Study on oil and gas accumulation conditions of Shawan Formation in Chepaizi Swell, Junggar Basin. *Xinjiang Oil Gas* **2020**, *16*, 1−7. (in Chinese with English Abstract)
(49) Tan, C. P.; Yu, X. H.; Li, S. L.; Xu, L.; Chen, B. T.; Li, S. L.; Shan, X. Sedimentary Characteristics of Fan Delta of the Badaowan Formation, Sikeshu Section, Southern Junggar Basin. *Geoscience* **2014**, *28*, 181−189. (in Chinese with English abstract)
(50) Song, J. Y. Geochemical Characteristics and Source Analysis of Crude Oil in Chunguang Block of Chepaizi Project in Junggar Basin. Master Degree Thesis; Yangtze University, 2018. (in Chinese with English abstract).
(51) Wang, Y.; Liu, L. F.; Ji, H. C.; Song, G. J.; Wang, X. M.; Sheng, Y. Structure of a pre-Triassic unconformity and its hydrocarbon transporting characteristics, Wuerhe-Fengnan area, Junggar Basin, China. *J. Pet. Sci. Eng.* **2018**, *173*, 820−834.
(52) Gao, X. Y.; Liu, L. F.; Wang, Y.; Jiang, Z. X.; Liu, G. D.; Shang, X. Q. The characteristics of unconformity surface at the bottom of the Paleogene and its significance in hydrocarbon migration in the Sikeshu Sag of the Junggar Basin, Northwest China. *Chin. J. Geochem.* **2013**, *32*, 181−190.
(53) Hu, H. W.; Zhang, Y. Y.; Guo, S. J.; Zhuo, Q. G. Origins and hydrocarbon accumulation significance of bitumen in the deeply buried Jurassic reservoirs in the southern Junggar Basin. *Acta Geol. Sin.* **2020**, *96*, 1883−1895.
(54) Liu, G.; Li, J. Z.; Qi, X. F.; Zhu, M.; Yuan, B.; Pang, Z. C. Reservoir formation process of the lower accumulation assemblage in the west part of the southern Junggar Basin: Case study of Well Dushan1 in the Dushanzi anticline. *Nat. Gas Geosci.* **2021**, *32*, 1009−1021.
(55) Chalmers, G. R. L.; Bustin, R. M. A multidisciplinary approach in determining the maceral (kerogen type) and mineralogical composition of Upper Cretaceous Eagle Ford Formation: Impact on pore development and pore size distribution. *Int. J. Coal Geol.* **2017**, *171*, 93−110.
(56) Lin, X. Y.; Qin, J.; Chen, Z.; Xu, Y.; Deng, X. H. Research on Evaluation and Distribution of Source Rocks in Sikeshu Sag of the Southern Junggar Basin. *J. Oil, Gas Technol.* **2013**, *35*, 1−5. (in Chinese with English abstract)
(57) Taylor, K. E. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.: Atmos.* **2001**, *106*, 7183−7192.
(58) Hintze, J. L.; Nelson, R. D. Violin plots: a box plot-density trace synergism. *Am. Stat.* **1998**, *52*, 181−184.
(59) Guangyou, Z. H. u.; Jin, Q.; Zhang, S. W.; Zhang, L. Y.; Guo, C. C. Salt lake-saline lake sedimentary combination and petroleum accumulation in the Bonan Sag. *Acta Mineral. Sin.* **2004**, *24*, 25−30.
(60) Qiu, N. S. Methods of thermal history reconstruction of sedimentary basins and their application in oil and gas exploration. *Mar. Orig. Petrol. Geol.* **2005**, *10*, 45−51.
(61) Peters, K. E.; Scott Ramos, L. S.; Zumberge, J. E.; Valin, Z. C.; Scotese, C. R.; Gautier, D. L. Circum-Arctic petroleum systems identified using decision-tree chemometrics. *AAPG Bull.* **2007**, *91*, 877−913.
(62) Xu, Y. D.; Chang, X. C.; Shi, B. B.; Wang, Y.; Li, Y. Geochemistry of severely biodegraded oils in the Carboniferous volcanic reservoir of the Chepaizi Uplift, Junggar basin, NW China. *Energy Explor. Exploit.* **2018**, *36*, 1461−1481.
(63) Peters, K. E.; Moldowan, J. M. *The Biomarker Guide, Interpreting Molecular Fossils in Petroleum and Ancient Sediments*; Prentice Hall: New Jersey, 1993.
(64) Seifert, W. K.; Michael Moldowan, J. M.; Demaison, G. J. Source correlation of biodegraded oils. *Org. Geochem.* **1984**, *6*, 633−643.

https://doi.org/10.1021/acsomega.2c07991
ACS Omega 2023, 8, 10314−10334

(65) Chung, H. M.; Rooncy, M. A.; Toon, M. B.; Claypool, G. E. Carbon isotope composition of marine crude oils. *AAPG Bull.* **1992**, *76*, 1000−1007.
(66) Kotsiantis, S. B. Supervised machine learning: a review of classification techniques. *Informatica* **2007**, *31*, 249−268.
(67) Zhang, L. P.; Bai, G. P.; Zhao, X. Z.; Zhou, L. H.; Zhou, S. S.; Jiang, W. Y.; Wang, Z. Y. Oil-source correlation in the slope of the qikou depression in the bohai bay basin with discriminant analysis. *Mar. Pet. Geol.* **2019**, *109*, 641−657.
(68) Li, M.; Larter, S. R.; Stoddart, D.; Bjoroey, M. Liquid chromatographic separation schemes for pyrrole and pyridine nitrogen aromatic heterocycle fractions from crude oils suitable for rapid characterization of geochemical samples. *Anal. Chem.* **1992**, *64*, 1337−1344.
(69) Tian, K.; Yin, J. Y.; Zeng, J. H. Hydrocarbon migration pathway based on quantitative grain fluorenscence techniques: A case of sandstone carrier beds of Shawan formation in Chepaizi Uplift. *J. Northeast Pet. Univ.* **2015**, *39*, 42−49. (in Chinese with English abstract)
(70) Jiang, Z. X.; Wang, X. D.; Pang, X. Q.; Zhang, J.; Tian, F. H. Paleohydrocarbon water contact restoration of typical Silurian oil and gas reservoirs in the northern Tarim Basin. *Earth Science—J. China Univ. Geosci.* **2006**, *31*, 201−208. (in Chinese with English abstract)
(71) Li, S. M.; Pang, X. Q.; Liu, K. Y.; Fu, X. L. A new approach to discriminate oil fluid inclusions quickly—Quantitative Grain Fluorescence (QGF) technique and its application. *Pet. Geol. Exp.* **2006**, *28*, 386−390. (in Chinese with English abstract)
(72) Liu, H. J.; Zhang, Z. H.; Qin, L. M.; Zhu, L.; Xi, W. J. Accumulation mechanism of light oils in Chepaizi uplift belt of Junggar Basin. *J. China Univ. Pet.* **2011**, *35*, 34−40. (in Chinese with English Abstract)
(73) Liu, H. J.; Duan, J. M.; Qin, L. M.; Zhang, Z. H.; Xi, W. J. Geochemical Characteristics and Sedimentary Environment Analysis of Cretaceous Source Rocks in Central and Southwestern Junggar Basin. *Geol. Sci. Technol. Inf.* **2013**, *32*, 25−30. (in Chinese with English Abstract)

https://doi.org/10.1021/acsomega.2c07991
ACS Omega 2023, 8, 10314−10334